«Una clase de introducción a la Física con un profesor estrella».
Scientific American

Las ideas fundamentales del Universo II

Nº1 best seller *The New York Times*

Física cuántica: cuantos y campos

Sean Carroll

Profesor de Física en el Instituto de Tecnología de California (Caltech)

arpa

LAS IDEAS FUNDAMENTALES DEL UNIVERSO II

Sean Carroll

LAS IDEAS FUNDAMENTALES DEL UNIVERSO II

Física cuántica: cuantos y campos

Traducción de Carmen Álvarez Benítez

arpa

Título original: *The Biggest Ideas in the Universe: Quanta and Fields*

© del texto: Sean Carroll, 2024
© de la traducción: Carmen Álvarez Benítez, 2024
© de esta edición: Arpa & Alfil Editores, S. L.

Primera edición: octubre de 2024

ISBN: 978-84-10313-35-4

Maquetación: Àngel Daniel Producción del ePub: booqlab Arpa
Manila, 65

08034 Barcelona

arpaeditores.com

Reservados todos los derechos.
Ninguna parte de esta publicación puede ser reproducida, almacenada o transmitida por ningún medio sin permiso del editor.

ÍNDICE

	Introducción
UNO	Funciones de onda
DOS	Medición
TRES	Entrelazamiento
CUATRO	Campos
CINCO	Interacciones
SEIS	Teoría de campo efectivo
SIETE	Escala
OCHO	Simetría
NUEVE	Teoría de campo de gauge
DIEZ	Fases
ONCE	Materia
DOCE	Átomos
	Apéndice: Transformadas de Fourier

A Ariel

INTRODUCCIÓN

La historia de la física ha sido testigo de numerosas ideas brillantes y transformadoras. Pero cuando se trata de cambios verdaderamente revolucionarios —cambios de paradigma que trastornan nuestra forma de pensar sobre la naturaleza de la realidad— solo ha habido dos realmente: la mecánica clásica a finales del siglo XVII y la mecánica cuántica a principios del siglo XX.

La mecánica clásica fue el tema de *Las ideas fundamentales del Universo: Espacio, tiempo y movimiento*, publicación en la que hicimos hincapié en la idea del paradigma newtoniano/laplaciano, efectuando todo el recorrido hasta llegar al espacio-tiempo y la relatividad. Ahora ha llegado el momento de que pasemos a la cuántica.

La mecánica cuántica explica, según nuestros conocimientos actuales, la forma en que funciona el mundo. Los primeros indicios de la necesidad de un cambio surgieron de los trabajos de Max Planck y Albert Einstein, que indicaron que la luz no era simplemente una onda, como habían pensado los físicos hasta entonces. En las circunstancias adecuadas, la luz se presenta en forma de partículas que ahora llamamos **fotones**. Estas partículas son un ejemplo de los **cuantos** del título del libro: cantidades en «paquetes» discretos de energía que resultan de las reglas de la mecánica cuántica. Pero se trata de algo más sutil. En distintas circunstancias, las cosas que consideramos partículas, como los electrones, protones y neutrones, presentan un comportamiento ondulatorio. La mecánica cuántica va a frustrar continuamente nuestro deseo de encuadrar el comportamiento de los sistemas físicos en casillas ordenadas y de sentido común.

No se sienta mal si las ideas de la mecánica cuántica le parecen extrañas al principio. La verdad es que ni los propios físicos se ponen

de acuerdo sobre lo que, en el fondo, dice la mecánica cuántica. Los físicos son extremadamente buenos a la hora de *utilizar* la mecánica cuántica. Podemos predecir la estructura de átomos y moléculas o calcular la dispersión de partículas entre sí con una precisión exquisita. Pero todo esto es una especie de caja negra. Los mejores físicos cuánticos del mundo no son unánimes sobre lo que ocurre para obtener los resultados que predicen y observan con tanto éxito.

Esta falta de consenso intelectual se debe a que la mecánica cuántica parece atribuir propiedades especiales al acto de «medir» u «observar» un sistema físico. En la física clásica, los objetos tienen propiedades como la posición y la velocidad, y se pueden medir directamente con tanta precisión como se quiera. Los objetos cuánticos parecen totalmente diferentes. Medir las propiedades de un sistema cuántico tiende a cambiarlas drásticamente. En algunas formas perfectamente razonables de concebir la teoría, una partícula como un electrón ni siquiera *tiene* propiedades como «posición» o «momento»: estas son posibles resultados de una medición, no características intrínsecas del propio sistema cuántico.

En la mayoría de los casos, no vamos a preocuparnos de nada de eso.[1] Aquí, en *Las ideas fundamentales del Universo*, nuestra actitud es la de físicos estrictos que utilizan ideas bien establecidas para hacer predicciones comprobables sobre el mundo. Con esto tendremos más que suficiente para reflexionar. Las cuestiones fundamentales son indiscutiblemente importantes. Comprender la mecánica cuántica a un nivel profundo fácilmente podría resultar crucial para ir más allá de nuestras teorías actuales y obtener una imagen mucho más completa de la realidad. Pero este libro se centrará en comprender los conceptos que subyacen a las teorías actuales y cómo nos han proporcionado una visión del mundo físico de una precisión sin precedentes.

Los conceptos en cuestión incluyen la propia mecánica cuántica, la teoría cuántica de campos, que surge de forma natural cuando se combina la mecánica cuántica con los requisitos de la relatividad especial, así como varias ideas profundas que han surgido dentro de la

teoría cuántica de campos, incluidos los diagramas de Feynman, la renormalización, las teorías de campo de gauge, la ruptura de la simetría y el teorema de conexión espín-estadística.

Se trata de una cantidad enorme de material, que me he esforzado por reducir a su esencia más pura. El truco, como con el libro anterior, es que vamos a incluir suficientes detalles matemáticos para entender las ideas de verdad, sin llegar a un nivel de detalle necesario para resolver problemas como lo haría un estudiante de posgrado que preparase su doctorado. Usted aprenderá las mismas ideas que ellos, pero no tendrá que pasar noches en vela resolviendo problemas.

Eso requiere una estrategia ligeramente distinta de la que empleamos en *Espacio, tiempo y movimiento*, aunque la aspiración básica es la misma. En ese libro, podría mostrarle literalmente todas las ecuaciones exactamente como las aprendería un profesional. En este caso, hay demasiada información para que eso sea factible. La teoría cuántica de campos está cargada de detalles puntillosos y abundante notación, que pueden impedir centrarse en las ideas principales. Y precisamente, son las ideas lo que nos importa. Así que habrá ocasiones en las que ignoraremos las constantes de acoplamiento, ocultaremos los índices y trataremos las matrices como números. Le prometo que todo esto es para ayudarle a entender lo que realmente ocurre, no para entorpecerlo.

Aun así, va a haber matemáticas. En *Espacio, tiempo y movimiento* introdujimos las ideas básicas del cálculo, incluidas las derivadas (tasas o velocidades de cambio) y las integrales (cantidades de cambio acumuladas). En capítulos posteriores nos ocupamos de los tensores, y del uso de letras griegas para denotar índices espacio-temporales. Todos estos conceptos van a estar presentes y vigentes aquí, así como las ideas físicas básicas de masa, energía y relatividad. Si ya está familiarizado con esos conceptos, el contenido de este libro será totalmente independiente. Si no es así, *Espacio, tiempo y movimiento* le transmitirá todo lo que necesita saber.

Será un viaje emocionante. En los albores del siglo XX, la mecánica clásica estaba firmemente arraigada. Veinticinco años más tarde, vimos las primeras formulaciones completas de la mecánica cuántica. Otros veinticinco años después, la electrodinámica cuántica constituyó la primera teoría cuántica de campos bien establecida. Y nuevamente veinticinco años después, los físicos habían elaborado el Modelo Estándar de la física de partículas, que sigue triunfando hoy en día. Este es el viaje en el que nos embarcamos, y que conlleva algunas de las ideas más asombrosas con las que el ser humano se ha topado jamás.

Una vez más, he tenido la extraordinaria suerte de recibir detalladas críticas constructivas sobre los borradores de este libro. Quisiera expresar mi enorme gratitud a Scott Aaronson, Justin Clarke-Doane, Ira Rothstein y Matt Strassler, que han conseguido que mi física sea clara y mis explicaciones no tan enrevesadas como hubieran sido de otro modo. Mi agente, Katinka Matson, me ha proporcionado sabios consejos a lo largo del proceso. Y un agradecimiento inmenso a mi editor, Stephen Morrow, que ha sido paciente y comprensivo y me ha ayudado especialmente a convertir esta serie de libros en algo que espero que resulte instructivo y haga disfrutar a los lectores.

Esta vez he tenido que ser especialmente paciente, ya que el proceso de escritura se vio interrumpido por una mudanza al otro extremo del país y la incorporación a la Universidad Johns Hopkins. Gracias a todos mis nuevos compañeros y alumnos por hacer que todo sucediera de la forma más fluida posible y por ser comprensivos cuando no siempre estuve disponible como me hubiera gustado.

Sobre todo, gracias a mi mujer, Jennifer, que hizo sus maletas y se mudó conmigo, asumió la mayor parte de la carga de dar forma a nuestro nuevo hogar y siempre mantiene la claridad de mis escritos. Espero con ilusión este nuevo y emocionante capítulo.

El gráfico del potencial de sombrero mexicano del capítulo 10 es una adaptación de un código que aparece en Mathematica, de Vitaliy Kaurov

(https://mathematica.stackexchange.com/questions/19578/how-can-i-make-a-plot-of-the-higgs-potential). La imagen del observatorio LIGO en Hanford del capítulo 11 procede de la colaboración científica LIGO (https://www.ligo.org/multimedia/gallery/lho-images/Aerial5.jpg). El gráfico de nucleidos del capítulo 12 se basa en un ejemplo de *Learning Scientific Programming with Python* [Aprendiendo programación científica con Python], de Christian Hill (https://scipython.com/).

---

1 Si le interesa, me permito recomendarle mi libro *Something Deeply Hidden* [Algo profundamente oculto], (Dutton, 2019), que presenta una introducción a los espinosos problemas intrínsecos a la mecánica cuántica. Y para una fascinante historia de este tema, eche un vistazo a *What Is Real?* [¿Qué es real?], de Adam Becker (Basic Books, 2018).

UNO

FUNCIONES DE ONDA

A medida que el siglo XIX llegaba a su fin, se habría perdonado a los físicos por creer que estaban en el camino correcto para entenderlo todo. El universo, según esta imagen poco definida, estaba hecho de partículas en movimiento empujadas por campos.

La idea de **campos** que llenaban el espacio había despegado realmente a lo largo del siglo XIX. Anteriormente, Isaac Newton había presentado una hermosa y convincente teoría del movimiento y la gravedad, y Pierre-Simon Laplace había demostrado cómo podíamos reformular esa teoría en términos de un campo gravitatorio que se extiende entre todos los objetos del universo. Un campo es simplemente algo que tiene un valor en cada punto del espacio. Este valor puede ser un simple número, un vector o algo más complejo, pero cualquier campo se extiende por todo el espacio.

Sin embargo, si lo único que importaba era la gravedad, el campo parecía algo opcional: un punto de vista que se podía adoptar o no, según las preferencias de cada uno. También era igualmente correcto pensar como Newton, directamente en términos de la fuerza creada sobre un objeto por la atracción gravitatoria de otros objetos sin que nada se extendiera entre ellos.

Eso cambió en el siglo XIX, cuando los físicos llegaron a pillarle el truco a la electricidad y el magnetismo. Los objetos cargados eléctricamente se ejercen fuerzas entre sí, lo que se atribuye de forma natural a la existencia de un campo eléctrico que se extiende entre ellos. Los experimentos de Michael Faraday demostraron que un imán en movimiento podía inducir corriente eléctrica al atravesar una espiral de alambre conductor sin llegar a tocarla, lo que apuntaba a la

existencia de un campo magnético independiente, y James Clerk Maxwell consiguió combinar estos dos tipos de campos en una única teoría del **electromagnetismo**, publicada en 1873. Esta unificación supuso un enorme triunfo, explicando un conjunto diverso de fenómenos eléctricos y magnéticos aplicando una teoría compacta. Las «ecuaciones de Maxwell» siguen dando quebraderos de cabeza a los estudiantes de física hoy en día.

Una de las exitosas implicaciones de la teoría de Maxwell fue la comprensión de la naturaleza de la **luz**. Más que un tipo de distinta sustancia, la luz es una **onda** que se propaga en los campos eléctrico y magnético, y también es conocida como **radiación electromagnética**. Pensamos en el electromagnetismo como una «fuerza», y lo es, pero Maxwell nos mostró que los campos portadores de fuerzas pueden vibrar y, en el caso de los campos eléctricos y magnéticos, esas vibraciones son lo que percibimos como luz. Los cuantos de luz son partículas llamadas fotones, así que a veces se afirma que «los fotones son portadores de la fuerza electromagnética». Pero, de momento, seguiremos pensando de forma clásica.

Supongamos que tenemos una sola partícula cargada, como un electrón. Si se deja aislada, tendrá un campo eléctrico a su alrededor, y habrá líneas de fuerza que apuntan hacia este electrón. La fuerza disminuirá siguiendo una ley cuadrática inversa, al igual que sucede con la gravedad newtoniana.[1]

Si movemos el electrón, ocurren dos cosas: en primer lugar, una carga en movimiento crea un campo magnético además de uno eléctrico. En segundo lugar, el campo eléctrico existente ajustará su orientación en el espacio, de forma que siga apuntando hacia la partícula. Y juntos, estos dos efectos (pequeño campo magnético, pequeña desviación en el campo eléctrico existente) presentan ondulaciones hacia el exterior, como las ondas que produce un guijarro arrojado a un estanque. Maxwell descubrió que la velocidad de estas ondulaciones es precisamente la velocidad de la luz, porque *son* luz. La luz, de cualquier longitud de onda, desde las ondas de

radio hasta los rayos X y gamma, es una vibración que se propaga en los campos eléctrico y magnético. Casi toda la luz que ve a su alrededor en este momento tiene su origen en una partícula cargada que se agita en algún lugar, ya sea en el filamento de una bombilla o en la superficie del sol.

Simultáneamente, en el siglo XIX, también se fue haciendo evidente el papel de las **partículas**. Los químicos, liderados por John Dalton, defendieron la idea de que la materia estaba formada por **átomos** individuales, con un tipo específico de átomo asociado a cada elemento químico. Los físicos se percataron más tarde, una vez que se dieron cuenta de que, al considerar los gases como conjuntos de átomos que rebotan, se podían explicar conceptos como la temperatura, la presión y la entropía.

Pero el término «átomo», tomado de la antigua idea griega de una unidad elemental indivisible de materia, resultó ser un poco prematuro. Aunque son los componentes básicos de los elementos químicos, los átomos de hoy en día no son indivisibles. Un resumen para salir del paso, pendiente de añadir detalles más tarde: los átomos están formados por un **núcleo** compuesto por **protones** y **neutrones** y rodeado de **electrones** en órbita. Los protones tienen carga eléctrica positiva, los neutrones carga cero y los electrones carga negativa. Podemos formar un átomo neutro si tenemos el mismo número de protones y electrones, ya que sus cargas eléctricas se anulan mutuamente. Los protones y los neutrones tienen aproximadamente la misma masa, siendo los neutrones tan solo un poco más pesados, pero los electrones son mucho más ligeros, con aproximadamente 1/1800 de la masa de un protón. Por tanto, la mayor parte de la masa de una persona o de otro objeto macroscópico se debe a los protones y neutrones. Los electrones, más ligeros, tienen más capacidad de movimiento y, por tanto, son los responsables de las reacciones químicas y del flujo de electricidad. Hoy en día sabemos que los propios protones y los neutrones están formados por partículas más

pequeñas llamadas **quarks**, que se mantienen unidas por **gluones**, pero a principios del siglo XX no se tenía ni idea de ello.

Esta imagen de los átomos se fue construyendo poco a poco. Los electrones fueron descubiertos en 1897 por el físico británico J. J. Thompson, que midió su carga y estableció que eran mucho más ligeros que los átomos. Así que, de algún modo, debía haber dos componentes en un átomo: los electrones, ligeros y con carga negativa, y una pieza más pesada y con carga positiva. Unos años más tarde, Thompson propuso una imagen en la que diminutos electrones flotaban dentro de un volumen mayor cargado positivamente. Esto se denominó el modelo del «pudin de ciruelas», en el que los electrones desempeñaban un papel similar al de las ciruelas.

El modelo del pudin de ciruelas no prosperó durante mucho tiempo. Un famoso experimento de Ernest Rutherford, Hans Geiger y Ernest Marsden consistió en disparar partículas alfa (que ahora se sabe que son núcleos de átomos de helio) sobre una fina lámina de pan de oro. Se esperaba que la mayor parte de las partículas pasaran directamente a través de ella, con sus trayectorias ligeramente desviadas en caso de que casualmente atravesaran un átomo e interaccionaran con los electrones (las ciruelas) o la masa difusa cargada positivamente (el pudin). Los electrones son demasiado ligeros para perturbar las trayectorias de las partículas alfa, y una carga positiva extendida sería demasiado difusa para tener mucho efecto. Pero lo que ocurrió fue que, aunque la mayoría de las partículas pasaron ciertamente sin ser afectadas, algunas rebotaron en ángulos enormes, incluso volviendo directamente hacia atrás. Esto solo podía ocurrir si había algo pesado y sustancial que hiciera a las partículas rebotar hacia atrás. En 1911, Rutherford explicó correctamente este resultado proponiendo que la carga positiva se concentraba en un núcleo central masivo. Cuando una partícula alfa entrante tenía la suerte de chocar directamente con el pequeño, pero pesado núcleo, se desviaba en un ángulo agudo, que es lo que se observó. En 1920 Rutherford propuso la existencia de protones (que eran simplemente

núcleos de hidrógeno, que ya habían sido descubiertos), y en 1921 teorizó la existencia de los neutrones (que finalmente fueron descubiertos en 1932).

Hasta aquí todo bien, piensa nuestro físico imaginario de *fin de siècle*. La materia está hecha de partículas, las partículas interaccionan mediante fuerzas, y esas fuerzas son soportadas por campos. Todo el mecanismo funcionaría según las reglas establecidas por el marco de la física clásica. En el caso de las partículas, este nos resulta bastante familiar: especificamos las posiciones y los momentos de todas las partículas. Luego utilizamos una de nuestras técnicas clásicas (leyes de Newton o su equivalente) para describir su dinámica. Los campos funcionan esencialmente de la misma manera, salvo que la «posición» de un campo es su valor en cada punto del espacio, y su «momento» es la velocidad a la que cambia en cada punto. En ambos casos se aplica la imagen clásica general.

La sospecha de que toda la física estaba a punto de resolverse era tentadora. Albert Michelson, en la inauguración de un nuevo laboratorio de física en la Universidad de Chicago en 1894, proclamó: «Parece probable que la mayoría de los grandes principios subyacentes [de la física] hayan sido firmemente establecidos».

Estaba muy equivocado.

Pero también estaba en minoría. Otros físicos, empezando por el propio Maxwell, admitieron que el comportamiento conocido de los conjuntos de partículas y ondas no siempre coincidía con nuestras expectativas clásicas. William Thomson, o Lord Kelvin, es a menudo víctima de una cita mal atribuida: «Ya no hay nada nuevo que descubrir en física. Todo lo que queda por hacer son mediciones cada vez más precisas». Su verdadera opinión era justo la contraria. En una conferencia pronunciada en 1900, Thomson destacó la presencia de dos «nubarrones» que se cernían sobre la física, uno de los cuales acabaría disipándose con la formulación de la teoría de la relatividad, y el otro, con la teoría de la mecánica cuántica.

RADIACIÓN DE CUERPO NEGRO

La historia de la ciencia es sutil y complicada, y el progreso rara vez sigue el camino recto que recordamos en retrospectiva. La mecánica cuántica, en particular, tuvo un desarrollo doloroso y desordenado. Vamos a pasar por alto muchos de los vericuetos históricos para centrarnos en dos fenómenos desconcertantes que desencadenaron la revolución cuántica: ondas con propiedades similares a las de las partículas y partículas con propiedades similares a las de las ondas.

Las propiedades de la luz como partícula se manifestaron en primer lugar. La idea surgió del estudio de la **radiación del cuerpo negro** (o «térmica»), que es la radiación emitida por un objeto que absorbe cualquier luz incidente, pero que, sin embargo, irradia solo porque tiene una temperatura distinta de cero. Para los físicos, la temperatura de un objeto caracteriza la agitación aleatoria de sus partículas constituyentes, y las partículas que se agitan aleatoriamente van a emitir radiación dependiendo de la velocidad a la que se muevan. Cuando observamos un cuadro, vemos una intrincada configuración de formas y colores que refleja la luz que incide sobre él. Un cuerpo negro, por el contrario, es lo que se obtiene cuando se apaga toda la luz ambiente y solo se deja que los objetos brillen debido a su temperatura; el resplandor de un elemento calefactor en una estufa eléctrica es un buen ejemplo. Todo lo que tiene una temperatura distinta de cero emite algo de radiación térmica, pero la radiación de cuerpo negro pura solo depende de la temperatura, sin verse afectada por el color, la reflectividad u otras propiedades del objeto. Un cuerpo negro a baja temperatura irradiará principalmente en longitudes de onda infrarrojas o incluso de radio y, a medida que aumentamos la temperatura veremos más luz visible, ultravioleta y, por último, rayos X.

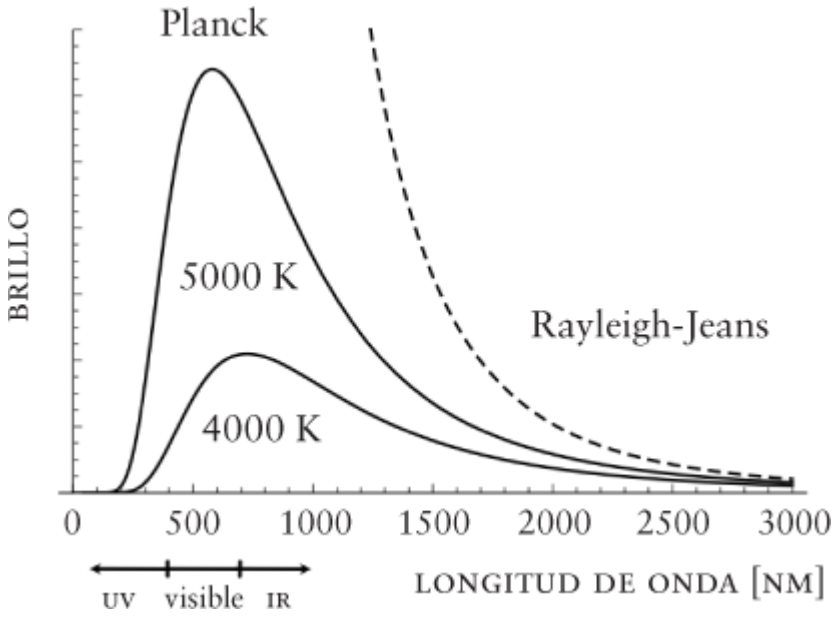
Así pues, la radiación de cuerpo negro representa un problema de física aparentemente sencillo (una vaca esférica, podríamos decir). Tiene una temperatura, y ninguna de sus demás propiedades importa.

La temperatura mide la energía cinética de los átomos del cuerpo que se agitan de un lado a otro, y esos átomos contienen partículas cargadas, por lo que esta agitación conduce a la emisión de radiación electromagnética. Nuestro problema físico es el siguiente: ¿cuánta radiación se emite en cada longitud de onda?

En el siglo XIX, los físicos se dedicaron a medir la radiación en función de la longitud de onda —el **espectro** del cuerpo negro— y a calcularla teóricamente. La curva medida era una belleza, ascendiendo desde cero a longitudes de onda cortas hasta un pico que dependía de la temperatura, y volviendo a decaer hasta cero a longitudes de onda largas.

Sin embargo, la situación teórica era muy confusa. Una teoría propuesta por Wilhelm Wien en 1896 parecía ajustarse bien a longitudes de onda cortas, pero divergía de los datos experimentales a longitudes de onda más largas. Otra, por John Strutt (Lord Rayleigh) en 1900, funcionaba al revés: se ajustaba bien a longitudes de onda largas, pero no a las cortas. De hecho, predecía una cantidad infinita de radiación en longitudes de onda cortas.

En general, se considera que el cálculo de Rayleigh, mejorado posteriormente por James Jeans, refleja con mayor exactitud lo que esperaríamos observar si el mundo hubiera sido realmente clásico. Su fracaso en longitudes de onda cortas ha sido bautizado como **catástrofe ultravioleta**, ya que los físicos se refieren de manera juguetona a todo lo que ocurre a corta distancia como ultravioleta, o simplemente «UV», y a todo lo que ocurre a larga distancia como infrarrojo o «IR». (Y a cualquier desajuste entre teoría y experimentación, como «catástrofe»). La relativa facilidad de comprensión de los fenómenos en la región IR, y la relativa dificultad de acertar con la región UV, volverán con sed de venganza cuando lleguemos a la teoría cuántica de campos.



El físico alemán Max Planck —a quien, según cuenta la historia, uno de sus profesores le dijo que no se dedicara a la física, porque «casi todo está ya descubierto»— decidió abordar el problema. En 1900 pudo escribir una fórmula que era un compromiso entre Wien y Rayleigh-Jeans, ajustando las observaciones tanto a longitudes de onda largas como cortas (así como entre ambas). Su resultado, la famosa **ley de radiación del cuerpo negro de Planck**, nos da el brillo B de un objeto a temperatura T en cada longitud de onda λ:

$$B(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1}. \qquad (1.1)$$

Aquí «exp» representa la función exponencial, $\exp(x) = e^x$. Además de la temperatura T y la longitud de onda l, esta expresión depende de tres parámetros fijos: la velocidad de la luz c, la constante de Boltzmann k_B de la termodinámica, y una nueva constante h que Planck tuvo que inventar para completar la fórmula. Ahora se conoce como la **constante de Planck**, número que aparece siempre allí donde la mecánica cuántica es relevante:

$$h = 6{,}626 \times 10^{-34} \text{ julios} \cdot \text{segundo}. \qquad (1.2)$$

Un julio es una unidad de energía, equivalente a la cantidad que consume una bombilla de un vatio en un segundo. Por diversas razones resulta que h aparece frecuentemente dividida por 2π, por lo que definimos la **constante de Planck reducida** como

$$\hbar = \frac{h}{2\pi}, \qquad (1.3)$$

y la pronunciamos como «h barra». Pronto nos daremos cuenta de que esta constante es tan ubicua que tendemos a elegir unidades en las que $\hbar = 1$, del mismo modo que es conveniente fijar la velocidad de la luz en $c = 1$ cuando trabajamos con relatividad. Pero, por ahora, mantengámosla al alcance.

Al principio, Planck no dedujo su fórmula, sino que más bien la adivinó. Elaboró las manipulaciones matemáticas correctas que combinarían los resultados de Wien y Rayleigh-Jeans en una única expresión compacta. Pero se esforzó mucho por encontrar una razón para que esa fórmula funcionara tan bien. Uno de los motivos por los que la tarea fue tan desafiante es que Planck era, en el fondo, un físico conservador. No le gustaba la mecánica estadística al estilo de Maxwell y Ludwig Boltzmann, la cual pretendía explicar las leyes de la termodinámica en términos del comportamiento colectivo de un gran número de átomos, e incluso era escéptico sobre la existencia de los propios átomos. Más tarde, siguió dudando de las principales ideas de la mecánica cuántica. Pero eso no le impidió dar con una ingeniosa hipótesis que resolvió el enigma de la radiación del cuerpo negro.

La conjetura crucial fue la siguiente: cuando una partícula cargada de un cuerpo se agita de un lado a otro y emite radiación electromagnética, la cantidad de energía contenida en esa radiación no puede ser un número cualquiera. Por el contrario, se emite en cantidades discretas que dependen de la frecuencia de la onda, la cual

está relacionada con su longitud de onda y la velocidad de la luz a través de la expresión $f = c/\lambda$. Las ondas luminosas de mayor frecuencia corresponden a «paquetes» de mayor energía según una fórmula ya famosa:

$$E = hf, \tag{1.4}$$

donde vemos de nuevo la aparición de la constante de Planck. Como a menudo es más cómodo utilizar la frecuencia angular $\omega = 2\pi f$ que la propia frecuencia, esta ecuación se suele escribir como $E = \hbar\omega$.

ONDAS QUE SE COMPORTAN COMO PARTÍCULAS

¿Por qué la energía de la radiación emitida se presenta en trozos discretos, en lugar de tener cualquier valor? Quizá porque las partículas del cuerpo negro solamente se agitan en cantidades discretas. Pero cabe pensar una explicación alternativa: no es que la agitación sea discreta, sino que la luz emitida lo es. En otras palabras, la luz es en realidad un flujo de entidades discretas: partículas.

Pero Planck no llegó a decir eso; su afirmación solo se refería a la cantidad de energía emitida, no a la forma que adoptaba. Que la luz sea una partícula o una onda es una cuestión antigua, que se remonta, por lo menos, a tiempos de Isaac Newton (que abogaba por las partículas) frente a Christiaan Huygens (que defendía las ondas). Una vez que aparecieron las ecuaciones de Maxwell en escena, los físicos quedaron bastante convencidos de que la luz era una onda. En particular, los experimentos habían demostrado que la luz interfiere consigo misma, como es de esperar cuando la parte positiva de una onda se cancela con la parte negativa. Por no hablar de que Maxwell había explicado la naturaleza de la luz en términos de su teoría del electromagnetismo, que tuvo éxito por muchas otras razones. Por

tanto, la idea de que la luz tuviera algo de partícula parecía improbable y los físicos dieron por superado ese concepto.

La persona que tuvo el valor de dar el salto fue un tal Albert Einstein de veintiséis años; 1905 se conoce como el *annus mirabilis* o «año milagroso» de Einstein. En una serie de artículos, formuló la relatividad especial, articuló la relación entre masa y energía y explicó el movimiento browniano (el movimiento aleatorio de partículas microscópicas en un líquido) en términos de colisiones atómicas, lo que ayudó a convencer a los científicos de la existencia de los átomos de una vez por todas. Cualquiera de estos logros habría dado renombre a un científico corriente, pero Einstein ni siquiera ganó el Premio Nobel por ninguno de ellos.

El que le valió el Nobel propuso la idea de los **cuantos de energía** de la luz. El «cuanto» (en latín, «*quantum*», o «*quanta*» en plural), se entiende como la «porción más pequeña en que puede subdividirse algo». El término **fotón** no se acuñó hasta más tarde, pero eso es precisamente lo que sugirió Einstein: que la luz está formada por partículas, ahora conocidas como fotones.

Lo hizo para explicar un fenómeno un tanto desconocido denominado **efecto fotoeléctrico**. Cuando la luz incide sobre un metal, a veces provoca la expulsión de un electrón energético. Este efecto no depende del brillo ni de la intensidad de la luz, como cabría esperar si la luz fuera una onda continua de energía, sino solo de su frecuencia. Eso tiene sentido si, en cambio, la luz se presenta en paquetes de cuantos (fotones) con energías individuales dadas por la ecuación (1.4). Cuando la energía de un fotón es suficiente para desprender un electrón, lo hace; cuando no lo es, no importa cuántos fotones se lancen, los electrones permanecerán seguros en su sitio. Esta imagen también basta, como señaló Einstein, para explicar la fórmula de Planck para la radiación de cuerpo negro.

Sin embargo, chocaba frontalmente contra todo lo que los físicos creían haber aprendido sobre la luz. La idea de que la luz es una onda no se había establecido por casualidad, sino que estaba respaldada por

evidencias experimentales y razonamientos teóricos de gran solidez. No se podía decir «la luz es una partícula, después de todo» y ya está. Más bien parecía que la luz se comportaba de forma ondulatoria la mayor parte del tiempo, pero que tenía propiedades de partícula en determinadas situaciones especiales. Si esto parece confuso y vago, es porque lo es. Tuvieron que pasar otras dos décadas dando tumbos antes de que empezara a surgir algo parecido a una explicación coherente. (E incluso ahora, un siglo después, no hay consenso sobre qué es lo que realmente ocurre).

PARTÍCULAS QUE SE COMPORTAN COMO ONDAS

Mientras tanto, Ernest Rutherford y sus colaboradores, entre los que destacaba el físico danés Niels Bohr, trabajaban para comprender la estructura atómica de la materia. Fueron los experimentos de Rutherford de 1911 los que habían demostrado que la mayor parte de la masa del átomo se concentraba en un núcleo denso y cargado positivamente situado en el centro. La cuestión era qué ocurría con los electrones, mucho más ligeros.

Una idea obvia sería que los electrones orbitaban alrededor del núcleo, de forma similar al movimiento de los planetas alrededor del Sol. El físico irlandés Joseph Larmor ya había sugerido antes algo parecido, y el físico japonés Hantaro Nagaoka propuso un marco alternativo en el que los electrones se moverían formando anillos (como los del planeta Saturno). El propio Rutherford era impreciso sobre lo que los electrones hacían, pero sabía que tenían que estar moviéndose alrededor del núcleo de alguna manera. Y conocía aproximadamente el tamaño de los átomos en su conjunto y, por tanto, de las órbitas de los electrones.

Y esto, como pronto se reconoció, suponía un enorme problema. Como acabamos de ver, las partículas cargadas en movimiento emiten radiación electromagnética. Este movimiento incluye el de los

electrones alrededor del núcleo atómico, al menos según las reglas de la mecánica clásica. Por tanto, nuestros electrones deberían emitir luz y, al hacerlo, perder energía. Como resultado, no deberían permanecer pacíficamente en órbita alrededor del núcleo. Un electrón en órbita se precipitaría rápidamente sobre el propio núcleo, proyectando la energía que pierde en forma de una sucesión de ondas electromagnéticas. Incluso se puede calcular cuánto tiempo tardaría, con una respuesta de aproximadamente 10^{-11} segundos. Pero si todos los átomos del universo duraran solo ese tiempo, alguien ya se habría dado cuenta.

Según las reglas de la mecánica clásica, en otras palabras, toda la materia hecha de átomos debería ser absolutamente inestable. Mesas, sillas, planetas y personas deberían colapsar y desintegrarse en una fracción de segundo. Pero eso no parece que ocurra. ¿Por qué?

Bohr propuso una primera respuesta en 1913. La gente conocía las ideas de Planck y Einstein sobre la radiación, así que la noción de «cuanto» ya estaba en el aire. Bohr sugirió que había algo vagamente cuántico en los electrones y su movimiento, al igual que en los fotones. Por supuesto, ya se pensaba en los electrones como partículas, pero la propuesta de Bohr consistía en que podían estar enlazados con los átomos únicamente en ciertas órbitas discretas, en lugar de mantener cualquier distancia. Los electrones ya son «cuantos» por el mero hecho de ser partículas, pero Bohr estaba diciendo que sus órbitas permitidas también están cuantizadas.

La idea funcionó bastante bien, al menos para el hidrógeno, el átomo más simple. Lo mejor de todo es que, para ajustarse a los datos, la condición de cuantización que Bohr tenía en mente equivalía a insistir en que el momento angular L de un electrón en órbita tendría que tener el valor

$$L = n\hbar\,, \tag{1.5}$$

donde n es cualquier número entero mayor que 0. Se trataba de una entrada en escena aparentemente milagrosa de la constante de Planck, propuesta inicialmente para comprender la radiación del cuerpo negro, en una fórmula para las órbitas de los electrones alrededor de los núcleos atómicos. Y la insistencia en un conjunto fijo de órbitas resolvió el problema de la estabilidad: los electrones podían, en el mejor de los casos, decaer hasta la órbita de menor energía, no hasta el fondo del todo donde está el núcleo. No hay ningún estado de energía más bajo al que puedan ir, lo que sugiere que los átomos deben ser estables una vez que todos sus electrones estén en sus órbitas de energía más baja.

El modelo de Bohr fue un avance importante, pero no explicaba *por qué* los electrones eran tan exigentes con sus órbitas. Finalmente, en 1924, el físico francés Louis de Broglie propuso una explicación como parte de su tesis doctoral. En retrospectiva, se trata de una idea sencilla: si la luz tiene propiedades similares a las de las partículas, ¿no deberíamos imaginar que las partículas también tienen propiedades similares a las de las ondas?

Seguía existiendo la cuestión de qué podría significar que algo tan manifiestamente parecido a una partícula como un electrón «tuviera propiedades ondulatorias». Parece que De Broglie tenía la sensación de que los electrones estaban formados tanto por partículas como por ondas, y que las partículas «perseguían» a las ondas. Pero la idea importante era imaginar que las ondas tenían una longitud de onda correspondiente, que podía relacionarse con el momento de la partícula, $p = mv$. Por lo tanto, la **longitud de onda de De Broglie** es

$$\lambda = h/p, \tag{1.6}$$

donde h es, una vez más, la constante de Planck. Esta longitud de onda podría utilizarse, según la idea de De Broglie, para predecir fenómenos similares a los ondulatorios, tales como la interferencia constructiva o destructiva al juntarse dos «ondas de materia».

Y lo que es más importante, el modelo de De Broglie sugería una explicación natural para las órbitas cuantizadas de los electrones de Bohr. A saber: las órbitas tienen que tener exactamente el tamaño adecuado para que quepan en ellas un número entero de longitudes de onda. En otras palabras, la onda tiene que volver al mismo valor del que partió cuando da una vuelta alrededor de la órbita. Esa condición resulta ser justamente la adecuada para reproducir las órbitas descritas por el modelo de Bohr, lo que sugiere claramente que esta idea iba por buen camino.

MECÁNICA CUÁNTICA

Pero los cuantos de luz de Einstein, las órbitas discretas de Bohr y las ondas de materia de De Broglie seguían pareciendo más un saco de ideas que una teoría completamente desarrollada. La primera formulación completa y rigurosa de lo que hoy llamamos **mecánica cuántica** fue propuesta en 1925 por un trío de físicos alemanes: Werner Heisenberg, Max Born y Pascual Jordan. La idea básica, que ahora se conoce como **mecánica matricial**, se le ocurrió por primera vez a Heisenberg mientras intentaba recuperarse de la fiebre del heno en una convalecencia en la isla de Helgoland. La idea era tan simple como revolucionaria: si las órbitas de los electrones son tan problemáticas, neguemos que exista el concepto de «órbitas de los electrones». Olvidémonos de lo que «realmente ocurre» y centrémonos solamente en lo que podemos *observar*.

Heisenberg propuso considerar el momento y la posición como **observables**, no como cantidades con valores definidos, sino como preguntas que podemos hacernos realizando mediciones. La pregunta «¿cuál es la posición del electrón?» no tiene respuesta antes de formularla. Es posible «medir la posición o el momento», pero al hacerlo se da existencia al resultado de la medición, no se revela una

verdad preexistente. Con esta idea, Heisenberg pudo deducir correctamente la forma en que los átomos emitían la luz.

Heisenberg aún era joven en aquel momento (veintitrés años), y le preocupaba que su modelo fuera demasiado audaz. Redactó un artículo en el que elaboraba su idea, pero antes de presentarlo lo envió a su experimentado colega Max Born, advirtiéndole de que «había escrito una locura». Born tenía la formación matemática adecuada para reconocer que el modelo de Heisenberg se expresaba mejor en el lenguaje de las **matrices**, sustituyendo las cantidades numéricas individuales por matrices cuadradas de números. Born y su antiguo alumno Pascual Jordan escribieron su propio artículo de seguimiento. A continuación, los tres colaboraron en otro trabajo más, en el que desarrollaron los detalles.

El problema era que, al igual que Heisenberg no estaba familiarizado con la idea de las matrices antes de que Born atara cabos, la mayoría de los físicos tampoco lo estaban. El formalismo matemático parecía bastante sólido, pero el significado físico subyacente era oscuro y la comunidad de físicos se mostraba reacia a declarar la victoria tan pronto.

Sin embargo, poco después, el físico austriaco Erwin Schrödinger propuso un enfoque aparentemente diferente. Siguiendo a De Broglie, Schrödinger volvió a situar las ondas en el centro de la historia, y su teoría recibió el nombre de **mecánica ondulatoria**. Con el tiempo se demostró que la mecánica matricial y la mecánica ondulatoria son dos formas equivalentes de representar la misma teoría física, por lo que hoy en día hablamos simplemente de **mecánica cuántica**.

Para ampliar la idea de las ondas de materia de De Broglie, Schrödinger propuso lo que ahora llamamos **función de onda**, a menudo escrita $\Psi(x)$, donde Ψ es la letra griega mayúscula psi —un nombre muy poco romántico para algo que llegaría a tener una importancia central para la naturaleza fundamental de la realidad—. Del mismo modo que la posición y el momento de cada componente especifican el estado clásico de un sistema, la función de onda

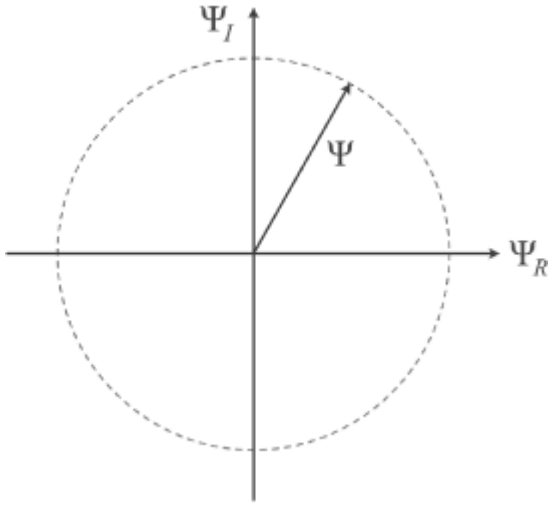
especifica el **estado cuántico** de un sistema. Si estamos considerando una sola partícula, su función de onda asigna un número a cada localización en el espacio, como ocurre con cualquier otro tipo de onda. Cuando tenemos más de una partícula, las cosas no son tan sencillas, porque la función de onda *no* es una onda cualquiera. Esto se debe al **entrelazamiento**, del que hablaremos largo y tendido en capítulos posteriores.

¿Qué representa físicamente la función de onda? Pues es una buena pregunta. Schrödinger la concibió inicialmente como algo bastante tangible, como la densidad de la materia. Pero al final se reinterpretó como una forma de calcular probabilidades de los resultados de mediciones, como veremos. Dejemos esta cuestión crucial a un lado por el momento.

Una gran diferencia entre la onda de materia de De Broglie y la función de onda de Schrödinger es que la función de onda en un punto cualquiera es un **número complejo**, formado a partir de la combinación de un número real y un número imaginario:

$$\Psi(x) = \Psi_R(x) + i\,\Psi_I(x). \qquad (1.7)$$

En este caso, $i = \sqrt{-1}$ es la «unidad imaginaria». Las funciones $\Psi_R(x)$ y $\Psi_I(x)$ son las partes real e imaginaria, respectivamente, de $\Psi(x)$. Observe que tanto $\Psi_R(x)$ como $\Psi_I(x)$ son a su vez números reales; se multiplica un número real por para obtener un número imaginario, que es lo que ocurre en el segundo término del lado derecho de la ecuación (1.7). Se puede pensar en la parte real y la parte imaginaria como dos ejes en un **plano complejo**, ilustrado en la figura siguiente.



¿Por qué la función de onda es compleja en lugar de real? En última instancia, es porque así es como funciona la naturaleza, en lugar de hacerlo de otra manera. Pero la naturaleza compleja de $\Psi(x)$ permite una característica interesante: la *forma* general de la función de onda (de, por ejemplo, un electrón en un átomo) puede mantenerse fija, mientras que la función evoluciona, simplemente compensando su parte real y su parte imaginaria entre sí. En la figura (en la que solamente estamos ilustrando el valor de Ψ en una x determinada, en lugar de la función completa), esto corresponde a girar Ψ en círculo manteniendo fija su longitud.

Sin embargo, esto hace que la función de onda parezca un poco abstracta. Algo como un campo eléctrico parece relativamente concreto. El campo eléctrico tiene un valor, representado por un pequeño vector, en cada punto del espacio. Incluso podemos medir ese valor colocando allí una partícula cargada y observando cómo el campo eléctrico la va empujando por el espacio. La función de onda de Schrödinger no parece tan *real*, ni literal ni matemáticamente. ¿Cómo se colaron los números complejos en nuestra mejor descripción de la realidad? Es posible que se pregunte: ¿qué clase de objeto se puede medir para obtener un resultado numérico imaginario? Pero los físicos están más que dispuestos a dejar de lado estos fastidiosos asuntos si con ello consiguen una teoría sencilla que

haga predicciones comprobables, como sin duda es el caso de la mecánica cuántica.

LA ECUACIÓN DE SCHRÖDINGER

Debemos tener presente que lo que hizo que la función de onda obtuviera un éxito inmediato entre los físicos fue la presencia de una ecuación dinámica que regía sus cambios a lo largo del tiempo. Conocida ahora como la **ecuación de Schrödinger**, adopta distintas formas específicas según el sistema de que se esté describiendo. (Igual que podemos introducir diferentes fuerzas en la segunda ley de Newton, $\vec{F} = m\vec{a}$). En su forma más general y abstracta, la ecuación de Schrödinger se ve así:

$$\widehat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t}. \tag{1.8}$$

Esto puede parecer un poco aterrador, pero es solamente porque los símbolos no son familiares. En realidad, la ecuación es extremadamente sencilla. Pasaremos por algunos de los detalles, pero al final todo lo que tiene que recordar es que la ecuación de Schrödinger nos dice cómo evoluciona la función de onda con el tiempo.

Volvemos a ver que aparecen la unidad imaginaria i y la constante de Planck reducida $\hbar$. El lado derecho es simplemente la derivada parcial de la función de onda Ψ con respecto al tiempo t. Recuerde que esta derivada parcial es una forma de decir «mantener todo lo demás fijo, y calcular la velocidad de cambio con respecto a t». Esto es lo que hace que la ecuación de Schrödinger, a diferencia de la relación de De Broglie (1.6), sea una relación *dinámica*: se especifica la función de onda en un instante de tiempo, y la ecuación determina lo que pasará a ser en el instante siguiente, así como en los instantes

subsiguientes. La ecuación de Schrödinger encaja perfectamente con el paradigma laplaciano de la mecánica clásica, en el que la información que especifica el estado se conserva a medida que el sistema evoluciona. (Como veremos, la función de onda parece evolucionar de un modo totalmente distinto cuando se mide el sistema en lugar de no interaccionar con él, y este es el origen de todo el misterio de la mecánica cuántica).

La parte complicada es el lado izquierdo, $\widehat{H}\Psi$. La letra H representa el **hamiltoniano**, que nos resulta familiar por nuestra investigación de la mecánica clásica en *Espacio, tiempo y movimiento*. Allí teníamos la posición x y el momento p como coordenadas en el espacio de fases, y el hamiltoniano $H(x, p)$ era simplemente la energía del sistema escrita en función de dichas coordenadas.

La situación cuántica es más complicada, pero de un modo divertido. El hamiltoniano ya no es una función del espacio de fases, sino un **operador**. Le ponemos un sombrerito y lo escribimos como $\widehat{H}$ para que no se nos olvide. Por «operador» entendemos un procedimiento matemático que toma una función y devuelve otra función. Para ser pedantes, el operador hamiltoniano es un mapeado de la función original Ψ para dar una nueva función, denominada $\widehat{H}\Psi$:

$$\widehat{H} : \Psi(x) \rightarrow \widehat{H}\Psi(x) . \qquad (1.9)$$

En mecánica clásica, el hamiltoniano es solamente la energía; en mecánica cuántica, es un operador que descompone la función de onda en varias partes, pregunta «¿cuánta energía hay en esta parte?» y luego suma los resultados para obtener una nueva función. Es bueno que el hamiltoniano, al actuar sobre la función de onda, nos dé otra función en lugar de simplemente un número, ya que queremos que sea proporcional a $\partial\Psi/\partial t$, que es ciertamente una función de x por sí misma. (La función de onda y la función que obtenemos al actuar el hamiltoniano sobre ella, también dependen del tiempo; sin embargo,

para simplificar las cosas, no necesitamos escribirlo explícitamente ahora mismo).

De modo que así es como debemos pensar acerca de los estados cuánticos y su evolución, en la idea de Schrödinger. Tenemos un sistema, como una sola partícula moviéndose en un potencial unidimensional, y tiene algunas coordenadas, como el número sencillo x. Empezamos en algún instante inicial con una función de onda de valores complejos que depende de las coordenadas, $\Psi(x)$. Entonces aplicamos un operador hamiltoniano, $\widehat{H}$, que actúa sobre la función de onda para dar una nueva función. El hamiltoniano que utilicemos realmente dependerá del tipo de sistema que estemos considerando y, en particular, del tipo de energía que tenga. Y esa nueva función nos da $i\hbar$ veces la derivada de Ψ con respecto al tiempo. La velocidad a la que evoluciona la función de onda depende de la energía del estado cuántico; los estados energéticos evolucionan más rápidamente, y los estados de menor energía, más lentamente.

UNA PARTÍCULA

Basta ya de cosas abstractas sin sentido. Bajemos a la Tierra y pensemos en lo que hace realmente el operador hamiltoniano.

Esto resulta ser una tarea ardua. Todo sistema está descrito por algún hamiltoniano, ya sea una sola partícula, el Modelo Estándar de la física de partículas o el universo en su conjunto. Pero el hamiltoniano será diferente en cada caso. Gran parte del trabajo de un físico teórico consiste en decidir cuál debe ser el hamiltoniano correcto para un sistema, ya que esa elección rige la dinámica del sistema. Desde la perspectiva de Schrödinger, «elegir un hamiltoniano» equivale a «elegir las leyes de la física». Es análogo a «elegir qué fuerzas actúan sobre un sistema» en mecánica clásica.

El propio Schrödinger no pensaba de forma tan abstracta, al menos al principio. Comenzaba, razonablemente, con un sistema

simple: una única partícula de masa m moviéndose a lo largo de una dimensión en un potencial $V(x)$. Literalmente, una pelota rodando por un paisaje lleno de colinas, aunque concebido desde el punto de vista de la mecánica cuántica. Pero ciñámonos a las partículas no relativistas (que se mueven lentamente en comparación con la velocidad de la luz), ya que resulta que la combinación de la relatividad con la mecánica cuántica requiere la teoría cuántica de campos para hacer las cosas bien.

Lo bueno es que, por lo general —o la mayoría de las veces en cualquier caso—, podemos obtener el operador hamiltoniano cuántico de un sistema si partimos del hamiltoniano clásico de toda la vida. Para una partícula no relativista, sabemos cuál sería el hamiltoniano clásico. Se trata de la energía cinética (escrita en términos del momento) más la energía potencial, por lo que

$$H(x,p)=\frac{p^2}{2m}+V(x). \tag{1.10}$$

Para convertir esto en un operador hamiltoniano cuántico, podemos ir parte por parte, tratando tanto x como p como operadores en sí mismos, que ahora denotamos $\hat{x}$ y $\hat{p}$. La operación de $\hat{x}$ es simplemente «multiplicar por x», lo que no es tan difícil. La parte complicada es el momento $\hat{p}$. En mecánica clásica, esta era una variable independiente que ayudaba a definir el estado. Para cualquier x, podíamos considerar tranquilamente estados con cualquier p.

Esto ya no es cierto en la mecánica cuántica. La función de onda $\Psi(x)$ define el estado cuántico completo; no hay dependencia adicional de p. En su lugar, el momento es ahora un operador, que es proporcional a la derivada parcial de Ψ con respecto a x:

$$\hat{p}=-i\hbar\frac{\partial}{\partial x}. \tag{1.11}$$

No tiene que preocuparse demasiado por saber de dónde viene esto o por qué es cierto. Por el momento, solamente debe fijarse en que el impulso está relacionado con la derivada espacial de la función de onda, que no es más que la pendiente de la curva. Una Ψ que serpentea suavemente se caracteriza por un momento bajo, mientras que una Ψ que oscila bruscamente tendrá un momento alto.

Ahora todo lo que tenemos que hacer es insertar (1.11) en la expresión clásica para el hamiltoniano (1.10) si deseamos obtener una fórmula para el hamiltoniano cuántico. Y eso nos da la ecuación de Schrödinger en la forma en que Schrödinger la escribió originalmente:

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+V(x)\right)\Psi(x,t)=i\hbar\frac{\partial}{\partial t}\Psi(x,t). \qquad (1.12)$$

La forma abstracta de la ecuación de Schrödinger, (1.8), tiene el tipo de belleza austera que desearíamos para una ley fundamental de la física. La forma explícita (1.12) es un poco más tosca, pero tenemos que prestar atención al detalle si queremos hacer predicciones experimentales. Cada año, en las universidades de todo el mundo, innumerables jóvenes estudiantes de física pasan la noche en vela resolviendo esta ecuación para diversas situaciones físicas.

EL OSCILADOR ARMÓNICO SIMPLE

Afortunadamente, no tendremos que dedicar tiempo a resolver nada explícitamente. Nos interesan menos las soluciones específicas de la ecuación de Schrödinger y más el principio general: existe una función de onda, y obedece a una ecuación dinámica definida, que establece su velocidad de cambio proporcional a la energía de la función de onda (en realidad, del operador hamiltoniano que actúa sobre ella).

Aun así, echar un vistazo a un tipo específico de solución puede ser bueno para el alma. ¿Y qué mejor ejemplo podemos considerar que

nuestro viejo amigo el oscilador armónico simple, con el que nos topamos en *Espacio, tiempo y movimiento*? Este ejemplo no se ha elegido solamente por simplicidad, sino que será de vital importancia cuando lleguemos a la teoría cuántica de campos. Empezaremos con funciones continuas y veremos la aparición de «cuantos», que se relacionarán directamente con el motivo por el que los campos cuantizados se parecen a las partículas.

Recordemos que el oscilador armónico simple está definido por una energía potencial que es cuadrática en la coordenada en cuestión. Podemos escribir esta como

$$V(x) = \frac{1}{2} m\omega^2 x^2, \tag{1.13}$$

donde m sigue siendo la masa de la partícula, ω es la frecuencia angular del oscilador y x es la coordenada. Estamos tentados de decir que «x es la posición», pero eso correspondería a la intuición clásica. La posición es un observable, y podríamos obtener cualquier respuesta particular x si la midiéramos, pero antes de hacer la medición no existe nada parecido a «la posición de la partícula».

Por supuesto, hay un número infinito de soluciones de la ecuación de Schrödinger en el potencial del oscilador armónico, ya que podemos simplemente empezar con cualquier $\Psi(x)$ que nos apetezca y luego utilizar la ecuación para determinar cómo evoluciona con el tiempo. (El paradigma laplaciano, que nos resulta familiar de la física clásica). Pero existen ciertas soluciones que resultan especialmente interesantes: las que mantienen su forma fija con el paso del tiempo. Estas son también soluciones para las que la energía tiene un valor definido en lugar de ser algo incierto. Por tanto, se denominan **autoestados energéticos**. Existe un estado de energía mínima, denominado **estado fundamental**, y además los **estados excitados** con la siguiente energía más alta, y la siguiente más alta después de esta, y

así sucesivamente. Si los etiquetamos con $n = \{0, 1, 2, ...\}$, las energías de estos estados adoptan la forma

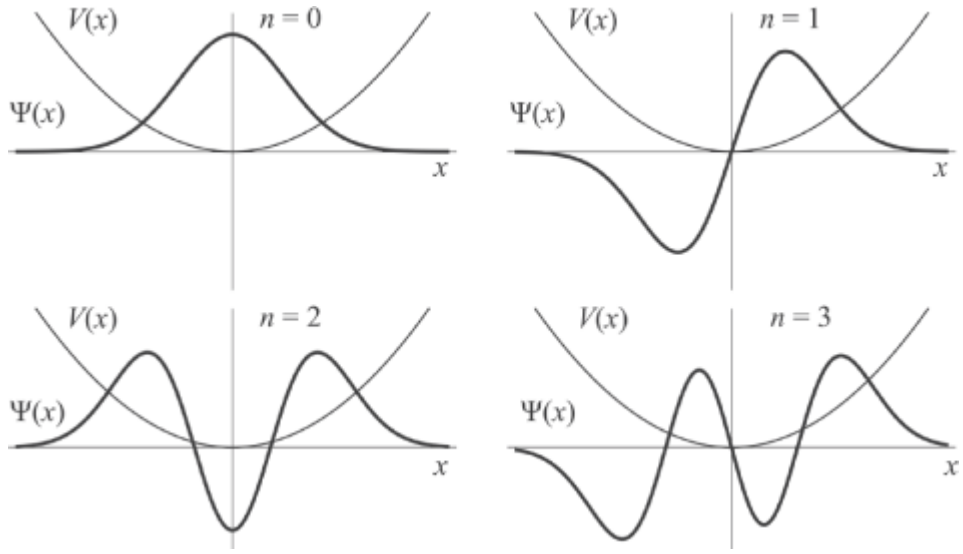
$$E_n = \hbar\,\omega\left(n+\frac{1}{2}\right), \tag{1.14}$$

donde w es la frecuencia angular del oscilador.

La forma matemática precisa de los autoestados de energía como funciones de x resulta ser complicada y no especialmente esclarecedora. Afortunadamente, podemos gesticular con las manos y hacer algunos dibujos para hacernos una idea intuitiva de lo que ocurre.

El hamiltoniano del oscilador armónico depende tanto de la energía potencial, que tiene su valor más bajo cerca de $x = 0$, como de la energía cinética, que tiene su valor más bajo cuando $\Psi(x)$ es una curva suave en lugar de oscilar fuertemente. Y, lo que es muy importante, la energía potencial se vuelve arbitrariamente grande en valores grandes de x, ya que crece como x^2. Así que cualquier estado de energía finita va a tener que acercarse a cero en $x = -\infty$ y $x = +\infty$. Si la función de onda se aproximara a cualquier valor fijo distinto de cero en $x = -\infty$, eso costaría infinita energía, ya que está multiplicando un potencial creciente en la ecuación de Schrödinger (1.12).

Así, viajando de $x = -\infty$ a $x = +\infty$, podemos más o menos adivinar cómo será la función de onda de menor energía ($n = 0$). Comenzará en cero cuando $x = -\infty$, pero no puede simplemente permanecer en cero; de lo contrario, no habría partícula en ningún sitio. Así que a medida que nos acercamos a $x = 0$, la función de onda va a aumentar suavemente, para luego empezar a reducirse de nuevo a cero. Nunca cruzará $\Psi = 0$, ya que estamos pensando en el estado de energía más bajo, y que va a variar lo más suavemente posible, sin ondulaciones extrañas.



La energía siguiente más alta (n = 1) se dispersará un poco más y tendrá un poco más de ondulación. De este modo, cruzará $\Psi = 0$ exactamente una vez: comenzando en cero, bajando en lugar de subir, volviendo a $\Psi = 0$ en $x = 0$, y luego subiendo y bajando de nuevo hasta cero. La segunda energía siguiente más alta (n = 2) cruzará dos veces: comenzando en cero, subiendo, luego volviendo hacia abajo a través del eje, luego hacia arriba de nuevo, y cayendo de nuevo a cero. El patrón se repite: cada autoestado de energía etiquetado n variará hacia arriba y hacia abajo, cruzando el eje $\Psi = 0$ exactamente n veces.

En medio de estos detalles matemáticos, es importante no pasar por alto que aquí se ha producido un pequeño milagro. Comenzamos nuestro viaje con la observación de Planck y Einstein de que había algo discreto, o «cuántico», en el comportamiento de los fotones, seguida de la aplicación de Bohr de una idea análoga a las órbitas de los electrones. Pero no hay nada discreto o cuántico en las funciones de onda o en la ecuación de Schrödinger. La propia función de onda es perfectamente continua, al igual que su evolución en el tiempo.

El oscilador armónico simple concilia esta tensión aparente: no es la función de onda o la ecuación que obedece lo que es discreto, es algún conjunto particular de *soluciones* de esa ecuación lo que tiene un carácter discreto. Es de ahí de donde proceden los cuantos.

Siempre que las funciones de onda estén inmovilizadas en el infinito, oscilarán alrededor de cero un número discreto de veces. Esto supone una analogía precisa a las cuerdas de un violín o una guitarra, que están literalmente atadas en cada extremo. Si pulsamos una cuerda de este tipo, puede vibrar con la mayor longitud de onda posible, que corresponde a la frecuencia fundamental, o con frecuencias discretas más altas, que corresponden a los sobretonos. La fórmula matemática será diferente para las funciones de onda, pero el mecanismo subyacente es el mismo.

Esto ocurre no solamente con el oscilador armónico, sino también con los electrones alrededor de los núcleos atómicos; sus niveles de energía se vuelven discretos debido al comportamiento de las soluciones apropiadas de la ecuación de Schrödinger, no porque haya algo fundamentalmente discreto en el espacio, el tiempo, la energía o cualquier otra cosa. Esta es también la razón por la que la teoría cuántica de campos, que es en última instancia una teoría de campos, es similar a una teoría de partículas: esos cuantos discretos de energía son las soluciones de la versión convenientemente adaptada de la ecuación de Schrödinger.

La principal ironía de la mecánica cuántica es que no hay nada fundamentalmente «cuántico» en ella. Vemos que en el universo ocurren ciertas cosas discretas porque así es como funcionan las soluciones de la ecuación de Schrödinger.

---

1 Más adelante en este libro veremos por qué tanto la gravedad como el electromagnetismo siguen leyes cuadráticas inversas. Se debe a que los campos subyacentes no tienen masa, lo que a su vez se debe a algo llamado «invariancia de gauge».

DOS
MEDICIÓN

El paradigma laplaciano de la mecánica clásica dice: dame el estado del sistema y las leyes de la física te dirán cómo evoluciona el sistema de un instante a otro. Si acaba de leer el capítulo anterior, se le puede perdonar si piensa que la mecánica cuántica sigue el mismo patrón básico. Claro, el «estado» es una función de onda en lugar de una colección de posiciones y momentos para algunas partículas, pero aun así se especifica en algún momento, y la ecuación de Schrödinger ayudará a hacerlo evolucionar a cualquier otro instante.

Esa no es toda la historia. Ni de lejos. Lo que diferencia radicalmente a la mecánica cuántica de la mecánica clásica es que hay momentos especiales en los que la regla dinámica central —la ecuación de Schrödinger— *no* parece describir cómo evoluciona el estado cuántico. (Tal vez lo haga en secreto, pero desde luego no lo *parece*). Esos momentos corresponden precisamente a cuando observamos, o medimos, alguna propiedad del sistema.

Esto parece, a primera vista, escandaloso. La mente lucha para comprender cómo el mero hecho de observar un sistema físico puede desempeñar un papel tan importante en su evolución. Se recurre a analogías clásicas para ayudar a dar sentido a una situación tan desconcertante. Tal vez sea como la gente que siempre pone una cara rara cuando intentas hacerles una foto. O tal vez sea simplemente que cada observación conlleva una interacción física de alguna clase; incluso cuando observamos las posiciones de los planetas en el cielo o las bolas de billar sobre una mesa, nuestra visión implica el intercambio de fotones, que llevan un poco de impulso, lo que seguramente afecta a la trayectoria del objeto que estamos mirando.

Estas analogías son malas. Lo que ocurre en la medición cuántica es algo mucho más profundo y no tiene un análogo directo en la física clásica. La medición es, en efecto, una interacción física, pero en la mecánica clásica esa interacción puede hacerse arbitrariamente pequeña, de modo que perturbe el sistema tan poco como queramos. En la mecánica cuántica, incluso la medición más leve y desapercibida puede cambiar drásticamente el estado cuántico.

El eje de la mecánica cuántica parece ser este: lo que *existe* no es lo que *se ve*. Están relacionados entre sí, por supuesto; no es que todo valga. Pero cuando medimos un sistema cuántico, no nos limitamos a observar —ni podemos hacerlo, en general— cuál era el estado cuántico antes de la medición. Observamos algún aspecto parcial e incompleto del estado, y en el proceso cambiamos el estado irrevocablemente.

Por eso, en la física cuántica, a diferencia de la clásica, la «medición» es una de las Ideas Fundamentales. Como ya se ha dicho, hay poco acuerdo sobre lo que ocurre entre bastidores. Físicos y filósofos han denominado esto el **problema de la medición**: ¿qué tienen de especial las mediciones, cuándo se producen precisamente y qué ocurre exactamente cuando hacemos una? Sin embargo, tenemos una idea lo suficientemente buena como para salir del paso acerca de las predicciones experimentales que nos permite hacer la mecánica cuántica, y en eso nos centraremos aquí.

COLAPSO DE LA FUNCIÓN DE ONDA

La introducción por parte de Schrödinger de la función de onda y de su ecuación correspondiente fue enseguida popular entre los físicos que se esforzaban por poner orden en el caótico lío de la teoría cuántica de mediados de la década de 1920. Pero acechaba un problema evidente: hay circunstancias muy claras en las que los electrones (y los fotones, para el caso) se comportan como partículas,

no como ondas. ¿Cómo conseguimos de la función de onda un comportamiento similar al de las partículas?

El propio Schrödinger tenía una esperanza: que cuando empezáramos con una función de onda algo extendida y la dejáramos evolucionar según su ecuación, la onda se localizara de forma natural alrededor de alguna posición real, posiblemente moviéndose a medida que pasara el tiempo. En otras palabras, las cosas podrían ser fundamentalmente ondulatorias, pero es posible que parecieran en su mayoría partículas porque la función de onda sería cero en casi todas partes, y distinta de cero solamente cerca de algún lugar en particular.

Esa esperanza no funcionó. Lo bello (y terrorífico) de una buena ecuación es que no hay margen de maniobra. La ecuación determina qué tipos de funciones pueden resolverla, y a esas soluciones no les importan las esperanzas ni los sueños de la gente. Resulta que las funciones de onda se comportan al revés de lo que quería Schrödinger: si empezamos con ellas relativamente localizadas, se extenderán y se volverán más difusas con el tiempo. A las funciones de onda les gusta ser como ondas, no como partículas, al menos en lo que respecta a la ecuación de Schrödinger.

Esto es un problema, si queremos explicar el mundo que vemos. Pensemos en un núcleo atómico radiactivo que se desintegra en partículas más ligeras. La ecuación de Schrödinger no deja lugar a dudas sobre el aspecto que tendrá la función de onda de las partículas recién creadas: se extenderá en una nube difusa, alejándose del núcleo original en forma aproximadamente esférica. Pero eso no es lo que vemos en realidad. Cuando un núcleo se desintegra, podemos tomar fotografías de las partículas emitidas, y describen trayectorias definidas, como cabría esperar de una partícula en movimiento. Vemos líneas, no esponjosas nubes. Este fue uno de los enigmas planteados por Albert Einstein en la famosa Quinta Conferencia Solvay de 1927, en la que los principales físicos del momento se reunieron en Bruselas (Bélgica) para codificar las nuevas reglas de la teoría cuántica.

Cuando enseñamos mecánica cuántica en los libros de texto modernos, la explicación de este fenómeno se articula en torno a la idea de que la toma de imágenes —ya sea realizada mediante una cámara de niebla en la década de 1920 o mediante un detector moderno en el Gran Colisionador de Hadrones a las afueras de Ginebra— es un acto de medición. Y la medición afecta a la función de onda de forma dramática.

En concreto, una medición cuántica es siempre la medición de alguna cantidad específica, el observable considerado. Una imagen del movimiento de una partícula recién creada es una medición de su posición en cada instante de tiempo. La regla es la siguiente: siempre que medimos un observable, cualquiera que fuera la función de onda antes de la medición, **colapsa** inmediatamente para dar algún valor definido de la cantidad observada. La nueva función de onda posterior al colapso evoluciona según la ecuación de Schrödinger, hasta que se observa y vuelve a colapsar.

He aquí cómo resolvemos el enigma de Einstein. En efecto, la función de onda inicial de una partícula creada a partir de un núcleo en desintegración se emite en forma esférica. Pero tan pronto como es detectada por algún proceso de observación, colapsa en un lugar concreto del espacio. Esta función de onda localizada también se extiende, pero en su mayor parte se aleja del núcleo original, ya que el momento se conserva. De modo que, cuando la observamos de nuevo poco después, y de nuevo un momento más tarde, estas observaciones se combinan para describir algo que se parece al rastro (línea) de una partícula en movimiento.

INDETERMINISMO CUÁNTICO

¿Cómo sabe la función de onda dónde localizarse cuando la observamos? No lo sabe. La posición en la que la función de onda colapsa implica un elemento irremediablemente aleatorio. Pero hay

una **probabilidad** definida para cada resultado de medición posible, y esa probabilidad está determinada por la función de onda.

La probabilidad se especifica mediante la **regla de Born**, propuesta por primera vez por Max Born en 1926. La regla de Born es bastante sencilla: la probabilidad de medir una posición x viene dada por el cuadrado de la función de onda asociada:

$$P(x) = |\Psi(x)|^2. \tag{2.1}$$

Únicamente tenemos que tener un poco de cuidado, porque $\Psi(x)$ es un número complejo, con partes real e imaginaria. Por eso ponemos las barras verticales alrededor de la función de onda en (2.1), para indicar que lo que realmente tenemos es el «módulo al cuadrado» de ese número complejo. Si descomponemos la función de onda en partes real e imaginaria, $\Psi = \Psi_R + i\,\Psi_I$, el módulo al cuadrado viene dado por la suma de los cuadrados de esas partes individuales:

$$|\Psi|^2 = \Psi_R^2 + \Psi_I^2. \tag{2.2}$$

El propio módulo $|\Psi| = \sqrt{|\Psi|^2}$ puede considerarse como la «longitud» del número complejo, que ilustramos en la figura del plano complejo del capítulo anterior. Así, donde la función de onda es grande, ya sea en su parte real o en su parte imaginaria, hay una probabilidad relativamente alta de ver la partícula allí si medimos su posición. Donde la función de onda es pequeña, hay una probabilidad relativamente baja de verla allí.

Pensemos por qué el módulo al cuadrado de la función de onda hace el trabajo que esperamos que haga una probabilidad. Las probabilidades de los sucesos tienen dos propiedades importantes: no pueden ser negativas, y cuando sumamos la probabilidad de cada posible resultado distinto tenemos que obtener 1. (Así es como

expresamos la idea de que algo inequívoco va a ocurrir, aunque no sepamos qué será).

La propiedad de no ser negativo es bastante evidente. El módulo al cuadrado es la suma de los cuadrados de dos números reales, lo que significa que no puede ser negativo. La propia función de onda no puede ser igual a la probabilidad, porque la función de onda puede ser negativa (o imaginaria). Pero el módulo al cuadrado resuelve el problema.

También queremos que la suma de todas las probabilidades sea 1. Pero como estamos midiendo la posición x de la partícula, no tenemos un conjunto discreto de resultados posibles, sino un continuo. No pasa nada; «sumamos» haciendo una integral:

$$\int P(x)\,dx = \int \left|\Psi(x)\right|^2\,dx = 1. \qquad (2.3)$$

En *Espacio, tiempo y movimiento* aprendimos que una integral no es más que la suma en todas las localizaciones del valor de lo que se está integrando. En castellano, esta expresión dice: «la probabilidad total de observar la partícula en todas las localizaciones posibles es igual a la unidad». La partícula tiene que ser observada en algún lugar, y no puede observarse en dos lugares a la vez. Decimos que una función de onda que satisface (2.3) está **normalizada**. Y si partimos de una función de onda normalizada y la dejamos evolucionar según la ecuación de Schrödinger, permanece normalizada: la probabilidad total de todos los resultados posibles sigue siendo 1.

Para que nos entendamos: según los libros de texto de mecánica cuántica, las funciones de onda pueden comportarse de dos maneras distintas. Cuando no se están midiendo, obedecen la ecuación de Schrödinger. Ese comportamiento no es muy diferente de lo que encontramos en la mecánica clásica: las funciones de onda evolucionan de forma suave, reversible (la información sobre el estado se conserva) y determinista. Pero en el momento de la medición, tiramos a Schrödinger por la ventana. La función de onda colapsa de

forma repentina, irreversible e indeterminista, de acuerdo con la regla de Born. No hay nada ni remotamente parecido en la física clásica.

Schrödinger, que razonablemente esperaba que su ecuación resolviera toda la historia, no estaba muy contento con la sugerencia de Born de que se utilizara para calcular probabilidades cuando las funciones de onda colapsan. «No me gusta —se quejó—, y lamento haber tenido algo que ver con ello».

Pero hubo alguna compensación (además del Premio Nobel, fama eterna y esas cosas). Si bien es cierto que la función de onda de una partícula que se desplaza libremente por el espacio tiende a dispersarse cada vez más, la velocidad de esa dispersión depende de la masa de la partícula. Las partículas más ligeras se dispersan rápidamente; las más pesadas lo hacen de manera más gradual. Por eso, cuando llegamos a las bolas de billar o a los planetas, la dispersión apenas es perceptible. Una vez localizados en el espacio, los objetos macroscópicos tienden a comportarse de forma bastante clásica. Nuestras viejas reglas newtonianas pueden recuperarse en las circunstancias adecuadas, como un caso límite del marco cuántico más general.

DUALIDAD ONDA-PARTÍCULA

En ninguna parte es más evidente la extraña dualidad de la evolución cuántica que en el célebre **experimento de la doble rendija**.

Ya en 1801, Thomas Young realizó un famoso experimento para demostrar la naturaleza ondulatoria de la luz. Esto fue antes de la unificación del electromagnetismo de Maxwell, por lo que todavía era plausible imaginar que la luz estaba hecha de partículas, y mucho antes de que Planck y Einstein reintrodujeran algo parecido a las partículas en la luz.

El experimento de Young aprovechó una propiedad crucial de las ondas: como oscilan hacia arriba y hacia abajo, pueden **interferir** entre sí. Si dos ondas se aproximan, habrá lugares en los que ambas vayan

hacia arriba o bien hacia abajo, y en estos casos unirán sus fuerzas para subir o bajar aún más («interferencia constructiva»). Pero, en otros lugares, una onda estará arriba y la otra abajo, por lo que se cancelarán entre sí («interferencia destructiva»). Las partículas no pueden interferir destructivamente de ese modo: lo único que se obtiene combinando partículas son más partículas.

Así que Young preparó un experimento en el que se apuntaba un haz de luz a dos finas rendijas, muy próximas entre sí. Y, efectivamente, cuando midió la luz en una placa detectora situada al otro lado de las rendijas, observó un patrón de interferencia de bandas claras y oscuras alternadas. La parte de la onda luminosa que atravesaba la rendija izquierda interfería con la parte que atravesaba la rendija derecha. Esto confirmó, según el pensamiento de la época, que la luz era una onda.

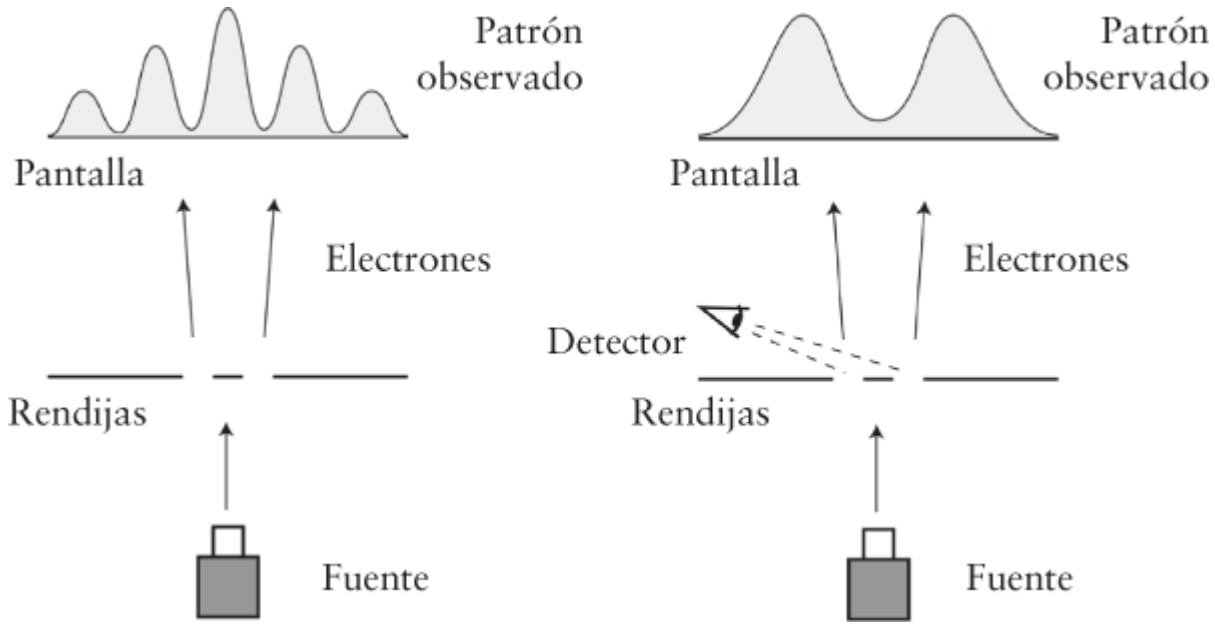
La versión cuántica es un poco más complicada. Imaginemos que dirigimos un solo electrón hacia dos rendijas situadas una cerca de la otra. Sabemos que el electrón tiene una función de onda, que atravesará las rendijas y presumiblemente interferirá consigo misma. Pero si colocamos un detector al otro lado, solamente observará una partícula, que aparece como un punto con una ubicación definida. ¿Qué significa entonces eso de que la función de onda interfiere consigo misma?

Para verlo, intentémoslo de nuevo con otro electrón y observemos su punto. Sigamos haciéndolo hasta que los puntos empiecen a acumularse para formar una imagen. Lo que vemos es exactamente el tipo de patrón de interferencia que esperamos de un comportamiento ondulatorio: bandas en las que se han acumulado muchos puntos, separadas por regiones oscuras en las que casi no hay puntos. Si los electrones fueran simplemente partículas clásicas, no habría interferencia; en su lugar, veríamos dos grupos de puntos, cada uno correspondiente a las partículas que pasan por cada rendija. Pero como cada electrón es una función ondulatoria hasta que se detecta en

la pantalla, vemos un patrón de bandas claras y oscuras que esperamos de la interferencia.

La cosa mejora. La regla es que las funciones de onda obedecen a la ecuación de Schrödinger, propagándose como una onda, hasta que son observadas, momento en el que colapsan. ¿Y si observamos «por qué rendija pasa el electrón»?

El patrón de interferencia desaparece. Si colocamos un detector que mida el electrón cuando atraviesa las rendijas, la función de onda del electrón colapsará en «pasar por la rendija izquierda» o «pasar por la rendija derecha». Entonces continuará hasta el detector y dejará un punto. Pero las posiciones de los puntos acumulados serán simplemente las de la rendija izquierda, más las de la rendija derecha. No hay interferencia destructiva, ya que las funciones de onda de cada electrón nunca pasaron por ambas rendijas a la vez.



Los debates sobre mecánica cuántica suelen incluir mucho discurso misterioso sobre la «dualidad onda-partícula». No hay razón para que sea confusa, al menos a nivel de lo que vemos al final. El electrón actúa como una onda cuando no es observado, y como una partícula

cuando lo es, porque la observación hace que la función de onda colapse en un lugar determinado.

MEDICIÓN Y REALIDAD

Sin embargo... *hay* algo confuso y misterioso, si queremos ahondar un poco más que «lo que vemos al final». Espero que la imagen anterior baste para convencerle de que sabemos utilizar la mecánica cuántica para casi todos los fines imaginables. Configure un sistema cuántico con alguna función de onda, déjelo evolucionar según la ecuación de Schrödinger y, a continuación, calcule la probabilidad de distintos resultados de mediciones mediante la regla de Born. Hay un poco de matemática complicada en el camino, pero el procedimiento en sí es lo suficientemente claro.

Lo que no está claro es lo que ocurre realmente. Es útil distinguir entre dos cuestiones sobre las que hay poco consenso entre los físicos cuánticos modernos: el problema de la medición y el problema de la realidad.

El **problema de la medición** es, en esencia, «¿qué ocurre realmente cuando medimos un sistema cuántico?». Al parecer, es algo importante, porque tenemos que inventar todo un nuevo conjunto de reglas para describir lo que ocurre cuando hacemos una medición. Pero hemos sido cautelosos sobre lo que realmente se considera una medición. ¿Se trata de cualquier tipo de interacción? ¿Tiene que haber intercambio de información? ¿Tiene que ser realizada por un agente consciente? ¿Qué ocurre si se observa un sistema sin prestarle realmente atención? ¿Cómo de rápido es el proceso de colapso de la función de onda, en cualquier caso?

A los físicos nos avergüenza o, al menos, debería avergonzarnos, no poder ponernos de acuerdo sobre las respuestas a estas preguntas. Hay respuestas plausibles, sin duda; el problema es que diferentes teorías responden a las preguntas anteriores de maneras muy distintas,

y no nos ponemos de acuerdo sobre cuál es la teoría correcta. Hablaremos algo más de las opciones en el próximo capítulo, una vez que hayamos reflexionado un poco sobre el entrelazamiento cuántico.

El **problema de la realidad** es sencillo: «¿cuál es la descripción cuántica de la realidad»? Esto no supone demasiada preocupación en la física clásica. (Los filósofos pueden preocuparse por cualquier cosa, pero las preocupaciones asociadas a la mecánica clásica parecen menos acuciantes). La realidad es un conjunto de partículas y campos que evolucionan en el espacio-tiempo siguiendo las leyes de la evolución clásica. Pero, en la teoría cuántica, donde utilizamos la función de onda para predecir las probabilidades de los resultados de las mediciones, las cosas no están tan claras. ¿Representa la función de onda, propiamente dicha, la realidad? ¿O no es más que una útil herramienta de cálculo? ¿Existen otras magnitudes físicas necesarias para describir la realidad en su totalidad? ¿O hablar de «realidad» es engañoso en cierto modo, y deberíamos contentarnos con poder predecir lo que vamos a medir?

Una vez más, no hay consenso entre los expertos. Algunos piensan que las funciones de onda capturan directamente la realidad; otros piensan que hay que aumentarlas con otras variables que describan cosas como la posición y el momento; también hay otros que no quieren hablar de la realidad en absoluto, o piensan que solamente hacemos que la realidad exista realizando mediciones.

No pasa nada: los científicos discrepamos constantemente y avanzamos reflexionando, recopilando nuevos datos y debatiéndolo todo. Pero es difícil explicar la mecánica cuántica de forma satisfactoria cuando ni siquiera nos ponemos de acuerdo sobre lo que dice la teoría.

Como ejemplo, consideremos de nuevo el experimento de la doble rendija. Cuando pensamos en un electrón como partícula, es tentador razonar en la línea de «el electrón pasa por una rendija o por la otra, simplemente no sabemos por cuál». Sin embargo, si la función de onda es la historia completa y definitiva, ese razonamiento es erróneo:

la función de onda del electrón pasa por ambas rendijas a la vez. No existe (bajo este punto de vista) algo así como «rendija por la que pasa el electrón». Del mismo modo, no existirían tales cosas como la posición ni el momento del electrón. No son cantidades que existan, sino simplemente resultados posibles de observaciones. Pero las observaciones no nos dan una visión pura de la realidad subyacente.

Al menos, eso es lo que diríamos si fuéramos «realistas» sobre las funciones de onda. Algunos no lo son. Las funciones de onda son confusas, y no nos ponemos de acuerdo, como comunidad (o especie), sobre lo que realmente son. De algún modo, somos perfectamente capaces de utilizarlas para hacer predicciones con una precisión asombrosa.

Pero necesitamos una forma de *hablar* sobre ellas, especialmente en un libro como este. Así que, aunque no nos detendremos aquí en cuestiones fundamentales, haremos la elección narrativa de hablar como si la función de onda fuera una representación directa y completa de la realidad. Esta elección entrará en juego cuando empecemos a hablar de campos cuánticos y partículas virtuales, cuya condición también plantea complicadas cuestiones filosóficas. Para nuestros propósitos, es la función de onda la que representa la realidad, y todo nuestro discurso sobre «partículas» y «campos» es un lenguaje útil para pensar en esa realidad en términos relativamente intuitivos. Si su formulación favorita de la mecánica cuántica prefiere otras formas de hablar, puede traducirla tranquilamente al lenguaje que le resulte más cómodo.

ESPACIO DE HILBERT

En la mecánica clásica, el concepto de **espacio de fases** —el conjunto de estados de un sistema clásico— desempeñaba un papel fundamental. Podría pensarse en el espacio de fases como el conjunto de todas las coordenadas y momentos posibles de cualquier sistema

imaginable. Una condición inicial consiste en un único punto en el espacio de fases; esta es toda la información que se necesita para resolver las ecuaciones clásicas y averiguar qué hará el sistema.

La idea análoga en mecánica cuántica es el **espacio de Hilbert**: el espacio de todos los estados cuánticos posibles. A diferencia de los estados clásicos, cuando tenemos dos estados cuánticos, podemos sumarlos o multiplicarlos por números (complejos), y obtener de este modo un estado cuántico diferente. Así, por ejemplo, dadas dos funciones de onda, $\Psi_1(x)$ y $\Psi_2(x)$, y dos números complejos, α y β, la cantidad

$$\Psi_3(x) = \alpha\Psi_1(x) + \beta\Psi_2(x) \tag{2.4}$$

es también una función de onda. (Si todo lo que hacemos es multiplicar una función de onda por un número, el estado resultante es físicamente equivalente al original).

No se pueden sumar las posiciones de dos partículas clásicas para obtener una tercera posición. Podría pensar que sí, pero eso es porque está pensando en coordenadas, y sumar coordenadas depende enormemente del sistema de coordenadas que se esté utilizando. Pero no hay ningún problema en sumar funciones de onda. Por tanto, estos estados cumplen los requisitos matemáticos para ser considerados «vectores» en algún sentido abstracto adecuado, por lo que el espacio de Hilbert es un ejemplo de **espacio vectorial**. Un espacio vectorial no es más que una colección de objetos (vectores) que pueden sumarse entre sí o cambiarse de escala multiplicándose por algún número. Así pues, hemos aprendido algo sobre los estados cuánticos: matemáticamente, son vectores, y la colección de todos esos vectores es el espacio de Hilbert.[1] El tipo de espacio de Hilbert que sea dependerá generalmente de la situación física, igual que el espacio de fases depende del sistema clásico que estemos considerando.

Puede parecer que estamos empleando de una forma poco familiar palabras que ya conoce. Ya hemos visto vectores: son flechitas, con

una longitud y una dirección. En los espacios vectoriales que hemos visto antes, podemos dibujar ejes y expresar el vector en términos de su componente a lo largo de cada eje. Cuánto del vector apunta en la dirección, cuánto en la dirección, y así sucesivamente. ¿Cómo encajan las funciones de onda en esta imagen tan intuitiva?

La respuesta es sutil, pero no descabellada. Las componentes de una función de onda $\Psi(x)$ son simplemente las **amplitudes** —los valores (complejos) de la función de onda— en cada punto x. Un momento, dirá: hay un número infinito de puntos etiquetados con x. Efectivamente, los hay. El espacio de Hilbert para una sola partícula es un espacio vectorial de *dimensión infinita*. También hay ejemplos de espacios de Hilbert de dimensiones finitas, pero siempre que empecemos con un sistema clásico que pudiera tener un número infinito de localizaciones (u otros posibles resultados de la observación), el correspondiente espacio cuántico de Hilbert es generalmente infinito-dimensional.

Estamos acostumbrados a pensar en espacios vectoriales bidimensionales, o tridimensionales, y quizá cuatridimensionales, una vez que nos convertimos en sofisticados habitantes del espacio-tiempo. El salto a un número infinito de dimensiones parece extremo y ciertamente es difícil de visualizar. El truco está en que no intentamos dibujar un espacio de Hilbert de infinitas dimensiones utilizando un conjunto infinito de ejes ortogonales. Simplemente pensamos en una componente para cada valor de la localización en el espacio x.

Paul Dirac, un brillante físico británico que fue uno de los pioneros de la mecánica cuántica, ideó una bonita notación para pensar en las funciones de onda como vectores: ponemos el símbolo del vector entre una barra vertical y un corchete angular, así: $|\Psi\rangle$. Esto se denomina **ket**, que proviene de la segunda mitad de la palabra «*bracket*»[2].

Esta notación nos ayuda a relacionar el espacio de Hilbert con el mundo de los vectores tal y como los hemos concebido hasta ahora. Dado un vector tridimensional ordinario $\vec{v}$, podemos expresarlo en términos de componentes como

$$\vec{v} = v_x\,\vec{e}_x + v_y\,\vec{e}_y + v_z\,\vec{e}_z, \tag{2.5}$$

donde $\vec{e}_x, \vec{e}_y, \vec{e}_z$ son vectores de base en las direcciones apropiadas. Podríamos traducir esto en notación de Dirac como $|v\rangle = v_x|e_x\rangle + v_y|e_y\rangle + v_z|e_z\rangle$.

La forma análoga de expresar el estado cuántico de una partícula en una dimensión espacial es

$$|\Psi\rangle = \int \Psi(x)|x\rangle\, dx. \tag{2.6}$$

Repasemos despacio esta expresión, ya que es un ejemplo paradigmático de cómo los profesores de física torturan a sus alumnos sobrecargándolos con pequeñas dosis de notaciones similares con significados ligeramente diferentes. A la izquierda tenemos $|\Psi\rangle$, que es el estado cuántico de nuestro sistema. Tiene existencia por derecho propio, independientemente de cómo se le represente, al igual que un vector tridimensional como $\vec{v}$ es algo en sí mismo, pero puede tener diferentes componentes dependiendo del sistema de coordenadas que utilicemos.

A la derecha tenemos una expresión específica para $|\Psi\rangle$, en términos de la posición x. Los vectores ordinarios pueden expresarse en términos de los ejes que se desee, y los estados cuánticos también podrán expresarse de diferentes maneras. Los kets $|x\rangle$ son los vectores de base, uno por cada valor de la localización en el espacio x, un número infinito de ellos en total. Al igual que escribimos el vector en (2.5) en términos de componentes v_i multiplicados por los vectores de base $\vec{e}_i$, sumados en las tres direcciones i, en (2.6) hemos escrito el estado cuántico como una suma de las amplitudes $\Psi(x)$ por los vectores de base $|x\rangle$, sumados para todos los valores de x. (Cuando empezamos con un signo integral y terminamos con dx, eso significa «suma la expresión encerrada para todos los valores de x»). Dado que la suma es para un número infinito de términos, la hemos escrito como una integral, pero el concepto subyacente es el mismo.

El hecho de que los estados cuánticos sean vectores es una afirmación profunda sobre la naturaleza de la realidad, no simplemente un exceso de formalismo. Los vectores de base que hemos elegido, $\{|x\rangle\}$, corresponden a algo observable: la posición de la partícula. Esto no es una coincidencia. De forma muy general, podemos encontrar una base para el espacio de Hilbert en términos de estados que tienen valores definidos para las cantidades observables. Aquí estamos considerando la posición, pero también podríamos utilizar el momento para definir una base diferente para el espacio de Hilbert.

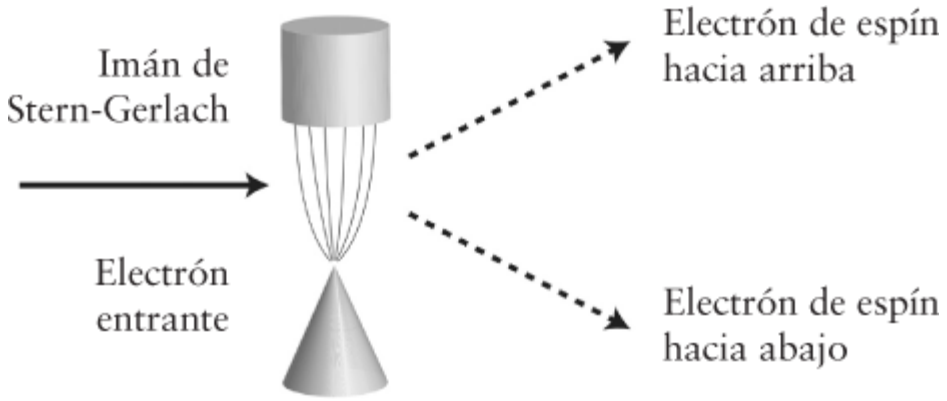
Si el estado cuántico de nuestra partícula coincidiera exactamente con algún vector de base $|x_*\rangle$, tendríamos la garantía de observar que la partícula se encuentra en la posición x_* si la midiéramos. En ese caso, podríamos decir que la partícula se encuentra en un estado de posición definida, o en un **autoestado de posición**, análogo a los autoestados de energía que consideramos previamente en el contexto del oscilador armónico simple.

Pero la mayoría de los estados no son de posición definida: son combinaciones de la forma (2.6). Así que, en esos casos, decimos que la partícula está en una **superposición** de muchas posiciones posibles. («Superposición» en el sentido de «múltiples posiciones, superpuestas», no «una posición realmente fantástica»). Esa es una característica clave de la mecánica cuántica, y una de las mayores motivaciones para gente (a diferencia de nosotros) a la que no le gusta pensar que el estado cuántico representa la realidad. Es natural asociar la «realidad» con lo que observamos. Cuando observo partículas, las veo en posiciones definidas, así que es tentador pensar que eso es lo que realmente existe. La mecánica cuántica parece decir algo profundamente distinto: que el estado de la partícula se encuentra en una superposición de muchos resultados de medida posibles antes de que la observemos.

CÚBITS

Una vez que usted (o von Neumann) se da cuenta de que los estados cuánticos pueden considerarse una especie de vectores, es natural preguntarse si el espacio de Hilbert en el que viven *tiene* que ser infinito. La respuesta es no: existen espacios de Hilbert bidimensionales, tridimensionales o con dimensiones de cualquier número entero mayor. (De nuevo, estas «dimensiones» son descripciones matemáticas, nada que ver con direcciones en el espacio físico tridimensional). Y hay un contexto natural en el que estos espacios de Hilbert finito-dimensionales son importantes: cuando hablamos del **espín** de una partícula cuántica.

En un famoso experimento sugerido por Otto Stern en 1921 y llevado a cabo por Walther Gerlach al año siguiente, se lanzó un haz de átomos de plata a través de un campo magnético colimado. Una partícula giratoria puede actuar como un pequeño imán, con un polo norte y un polo sur, y se desviará en un campo magnético dependiendo de cómo esté orientado el eje del imán. Cabría esperar que los átomos se desviaran varios grados, dependiendo de las orientaciones de sus espines. Lo que realmente vemos en el experimento de Stern-Gerlach es diferente: los átomos se desvían hacia arriba o bien hacia abajo. No hay ningún comportamiento intermedio.



La razón, como ahora sabemos, es la mecánica cuántica. Cada átomo tiene un estado cuántico que depende no solamente de su posición, sino también de su espín. Cuando lanzamos los átomos a través del campo y observamos cómo se desvían, estamos midiendo la magnitud de su giro o espín[3] a lo largo de un eje concreto. Y en el caso de los átomos de plata (así como de muchas partículas elementales, incluidos electrones, neutrinos y quarks) hay solamente dos resultados de medición posibles: «espín hacia arriba» o «espín hacia abajo». El momento angular de una partícula con espín hacia arriba es $+\hbar/2$ y el de una partícula con espín hacia abajo es $-\hbar/2$, donde $\hbar$ es nuestra vieja amiga, la constante de Planck reducida. Cuando estas son las dos únicas posibilidades, decimos que tenemos entre manos una **partícula de espín ½**.

Cuando estamos describiendo la posición de una partícula en una dimensión, hay un número infinito de posibles posiciones x que podríamos observar, y los estados cuánticos correspondientes viven en un espacio de Hilbert de dimensión infinita, con vectores de base $\{|x\rangle\}$. Cuando describimos el espín de una partícula de espín ½, hay dos posibles resultados de medición y, en consecuencia, un espacio de Hilbert bidimensional para describir los estados cuánticos. Si medimos el espín en la dirección del eje z, por ejemplo, podemos etiquetar los vectores de base como $|+z\rangle$ para espín hacia arriba y $|-z\rangle$ para espín hacia abajo, y escribir el estado como $|\Psi\rangle = \alpha|+z\rangle+\beta|-z\rangle$, donde α y β son números complejos que satisfacen $|\alpha|^2+|\beta|^2 = 1$. Los vectores de base son «autoestados de espín» (estados cuánticos con un valor de espín definido, no una superposición de múltiples posibilidades) en la dirección z.

También somos libres de medir el espín en la dirección del eje x, o del eje y, o de cualquier otro eje que queramos. Una vez más, solamente tendremos dos resultados posibles: espín hacia arriba o espín hacia abajo con respecto al eje que utilicemos. Un estado de espín hacia arriba en la dirección del eje x es $|+x\rangle$, espín hacia arriba en la dirección del eje y es $|+y\rangle$, y así sucesivamente.

Una vez que hemos especificado el estado cuántico en alguna base, escribirlo en una base diferente no requiere nueva información; es solamente un reempaquetado de la información existente que especifica el estado. En el caso del espín, esto puede verse porque los vectores de base con respecto al eje x o al eje y son solamente combinaciones de los vectores de base del eje z:

$$
\begin{aligned}
|+x\rangle &= \frac{1}{\sqrt{2}}|+z\rangle + \frac{1}{\sqrt{2}}|-z\rangle, \\
|-x\rangle &= \frac{1}{\sqrt{2}}|+z\rangle - \frac{1}{\sqrt{2}}|-z\rangle, \\
|+y\rangle &= \frac{1}{\sqrt{2}}|+z\rangle + \frac{i}{\sqrt{2}}|-z\rangle, \\
|-y\rangle &= \frac{1}{\sqrt{2}}|+z\rangle - \frac{i}{\sqrt{2}}|-z\rangle.
\end{aligned}
\tag{2.7}
$$

También puede sumar y restar combinaciones de estas ecuaciones para resolver igualmente $|\pm z\rangle$ en términos de $|\pm x\rangle$ o $|\pm y\rangle$. Y, a continuación, puede tomar el estado expresado en cualquier base particular y sustituir estas relaciones para convertirlo a cualquier otra base.

Estas sencillas ecuaciones constituyen el fundamento del célebre **principio de incertidumbre de Heisenberg**, como veremos en un momento. Nos recuerdan que es el estado lo que es relevante, y no la forma en que decidamos representarlo. La primera línea de la ecuación (2.7), por ejemplo, indica que un espín en un estado definido $|+x\rangle$ también puede considerarse, con la misma rigurosidad, situado en una cierta superposición de los estados $|+z\rangle$ y $|–z\rangle$. No hay nada parecido en la física clásica.

En el mundo clásico, la unidad básica de información que podemos contemplar es el **bit**: un número que solamente tiene dos valores, tradicionalmente representados como 0 o 1. En el mundo cuántico, la unidad básica de información se denomina, naturalmente, **cúbit o** «bit cuántico». En lugar de ser 0 o 1, un cúbit se representa mediante una superposición de dos vectores de base en un espacio de Hilbert bidimensional, como en

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle. \qquad (2.8)$$

Los coeficientes complejos a y b se denominan amplitudes correspondientes a $|0\rangle$ y $|1\rangle$, respectivamente. Del mismo modo, cualquier valor particular $\Psi(x)$ de la función de onda para una única partícula en la posición x es la amplitud para que la partícula se encuentre en esa posición.

Un bit de información es un concepto abstracto y puede representarse de diversas formas en un ordenador físico, desde dos posiciones de un interruptor hasta dos cantidades distintas de corriente en un circuito. Del mismo modo, un cúbit es un estado abstracto en un espacio vectorial bidimensional, que puede representarse mediante varios sistemas físicos diferentes. El espín de una partícula de espín ½ con respecto a un eje dado es una de esas representaciones. La interpretación física de lo que entendemos por los estados $|0\rangle$ y $|1\rangle$ dependerá de la tecnología que utilicemos para construir nuestros cúbits. La investigación moderna en computación cuántica ha explorado diversas formas de codificar cúbits en sistemas físicos.

REPASO DEL MOMENTO

Hay un curioso desequilibrio en la forma en que hemos estado hablando de las funciones de onda. En la mecánica clásica, el estado

de una partícula consiste tanto en su posición como en su momento, y esas dos cantidades son completamente independientes entre sí. Si le indico la posición en un instante de tiempo, no tiene ni idea de cuál es el momento. Pero hemos presentado la función de onda como $\Psi(x)$: algo que depende solamente de la posición. ¿Qué pasó con el momento?

No se preocupe, sigue ahí. Pero la posición y el momento no son variables independientes como lo eran en la mecánica clásica. Ambas son observables, y podemos calcular la probabilidad de observar valores diferentes cuando realizamos una medición. Y calculamos la probabilidad para ambas a partir del estado cuántico único.

El aparente desequilibrio se debe enteramente a la forma en que hemos elegido introducir la función de onda, como un número complejo asociado a cada posición posible en la que podríamos medir que estuviera la partícula. Hay otra forma de expresar el estado cuántico que es igual de válida: como un número complejo para cada *momento* posible que pudiéramos medir que tuviera la partícula. Si el momento se denota p, entonces podríamos escribir esta forma alternativa de la función de onda como $\tilde{\Psi}(p)$, donde la tilde (~) sobre la Ψ nos recuerda que es una función del momento, no de la posición, aunque represente el mismo estado físico cuántico. Entonces la probabilidad de encontrar un cierto valor de momento, si lo midiéramos, responde simplemente a la regla de Born, como de costumbre:

$$P(p) = \left|\tilde{\Psi}(p)\right|^2. \tag{2.9}$$

La información contenida en $\tilde{\Psi}(p)$ es precisamente la misma que la contenida en la función de onda original $\Psi(x)$. Es literalmente un cambio de base, exactamente igual que podemos cambiar la base de un espín de $\{|+z\rangle, |-z\rangle\}$ a $\{|+x\rangle, |-x\rangle\}$. En la **base del momento**, el equivalente de la ecuación (2.6), que expresa $|\Psi\rangle$ como una suma

extendida a todos los estados de posición, es expresarla como una suma para todos los estados de momento,

$$|\Psi\rangle = \int \tilde{\Psi}(p)\,|p\rangle\, dp, \qquad (2.10)$$

donde $|p\rangle$ es un vector de base con momento definido p (un autoestado del momento). Nótese que no hay tilde sobre la Ψ en el ket de la izquierda. Eso es porque se trata del *mismo* estado cuántico, no de uno nuevo. Lo único nuevo es la forma en que lo hemos representado como una suma extendida a todos los componentes, en este caso, los valores de la función de onda del momento $\tilde{\Psi}(p)$.

Las fórmulas explícitas para llevar a cabo la transformación de la base de posiciones a la base de momentos y viceversa son un poco complicadas, por lo que las hemos relegado al apéndice. El truco matemático pertinente se denomina **transformada de Fourier**, que es una forma de representar funciones como una suma de ondas sinusoidales en lugar de como valores explícitos en cada ubicación. La función de onda en la base de momentos $\tilde{\Psi}(p)$ es la transformada de Fourier de la función de onda en la base de posiciones $\Psi(x)$. Pero podemos ciertamente intuir lo que ocurre sin perdernos en demasiados detalles.

Una onda de sonido que incide en su tímpano puede considerarse una oscilación en movimiento de la densidad y la presión del aire que le rodea. Pero cuando escuchamos un piano u otro instrumento musical, nuestro cerebro no percibe todas las variaciones de la presión del aire a medida que aumenta o disminuye. En su lugar, oímos notas musicales —vibraciones a frecuencias específicas—. Dependiendo de nuestros conocimientos musicales, podemos incluso distinguir las distintas notas que se combinan para crear un acorde. Del mismo modo, cuando sintonizamos una emisora de radio, le estamos diciendo al aparato que se centre en una longitud de onda concreta: la de las

denominadas ondas de radio. Hay muchas otras ondas, solamente que las ignoramos.

Este es un ejemplo de una idea general que resulta útil siempre que consideramos fenómenos ondulatorios, ya sea en forma de ondas sonoras o de función de onda cuántica. Podemos pensar en esa onda como una función del espacio, que oscila arriba y abajo, *o bien* podemos considerarla definida por una determinada frecuencia o combinación de frecuencias. Así pues, dada cualquier función de onda $\Psi(x)$, podemos expresarla como una suma ponderada de ondas de cada longitud de onda posible. Eso es precisamente lo que consigue la transformada de Fourier.

La relación con el momento es sencilla: una función de onda que se parece a una onda sinusoidal como función de la posición es un estado con un valor definido de momento —un autoestado de momento—. La relación entre la longitud de onda y el momento correspondiente es precisamente la ecuación de De Broglie (1.6):

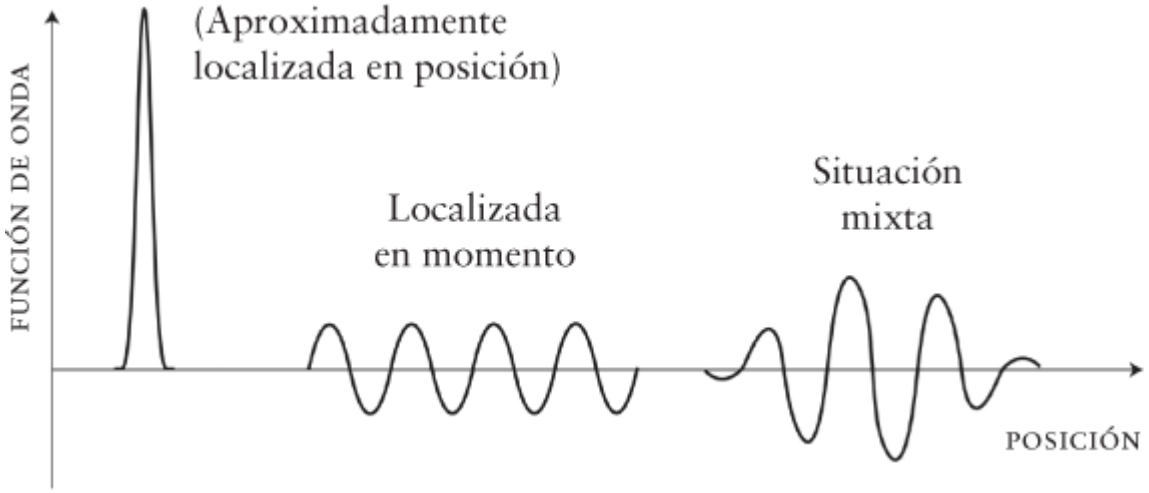
$$\lambda = h/p. \tag{2.11}$$

Así, los estados cuánticos con momento *elevado* se describen mediante funciones de onda sinusoidales con longitud de onda *pequeña*. Esta relación inversa entre tamaño y momento (y, por tanto, energía) desempeñará un papel importantísimo en la teoría cuántica de campos.

EL PRINCIPIO DE INCERTIDUMBRE

Consideremos una función de onda con momento definido. Como función de x, oscila hacia arriba y hacia abajo con una amplitud fija y exactamente la misma longitud de onda para todos los valores de x. Una función de onda de este tipo probablemente no pueda ser normalizada como en la ecuación (2.3) —la integral sería infinita, ya

que la función oscila eternamente en lugar de desvanecerse—, pero ese es solo uno de los molestos detalles matemáticos que decidimos pasar por alto para llegar a lo importante. Lo que sí podemos decir es que la probabilidad de observar la partícula en una posición determinada está muy dispersa por todas partes. Cuando el momento se conoce con precisión, no tenemos ni idea de dónde se verá la partícula.



No se trata solamente de una peculiaridad de estos estados concretos, sino de un ejemplo de una regla general: el principio de incertidumbre de Heisenberg. Podemos pensar en estados con momento definido, pero no tendremos ni idea de dónde se verá la partícula. En cambio, un estado con posición definida corresponde a una superposición de todos los momentos posibles, por lo que no tenemos ninguna idea de qué momento observaríamos. También existen estados de «paquete de ondas», con cierta dispersión tanto en x como en p, en la que ninguno de los dos está determinado con precisión.

Según el principio de incertidumbre, en cualquier estado cuántico concebible existe una incertidumbre mínima irreducible asociada a la combinación de posición y momento, de la forma:

$$\Delta x \, \Delta p \geq \frac{\hbar}{2} . \qquad (2.12)$$

Aquí Δx es la desviación esperada en los posibles resultados de una medición de posición, y Δp es la desviación esperada en los posibles resultados de una medición de momento. Pero la esencia del principio de incertidumbre no tiene nada que ver con las mediciones. Tiene implicaciones para las mediciones —si medimos la posición o bien el momento de forma precisa, la función de onda colapsará y no tendremos ni idea de cuál sería la otra magnitud inmediatamente después—, pero en realidad es una característica de los estados cuánticos incluso antes de que los midamos.

La cuestión no es que inevitablemente nos choquemos con un sistema cuántico mientras lo medimos y, por tanto, lo cambiemos. Se trata de que no existen estados en los que tanto la posición como el momento estén muy localizados al mismo tiempo. Esto es difícil de interiorizar si nos quedamos atascados en la intuición clásica, pensando en la posición y el momento como cosas que realmente existen; es más fácil de digerir si pensamos en ellos como conjuntos de posibles resultados de observación que derivamos a partir de un estado cuántico subyacente.

Los físicos opinan a veces que el principio de incertidumbre es bastante trivial una vez que se entiende lo que ocurre. Al fin y al cabo, es solamente un cambio de base. Podemos escribir el estado cuántico como una suma extendida a todos los estados de posición definida, o como una suma extendida a todos los estados de momento definido, pero un estado que es definido en posición es una superposición de estados de momento, y viceversa. Así que, por supuesto, es imposible tener un estado que sea definido en ambos. Es como el cambio de base de los espines, del eje z al eje x (o el que sea) como en la ecuación (2.7). Un estado de espín en x definido es la suma de los espines en z hacia arriba y hacia abajo, y viceversa. Un vector no puede apuntar con precisión a lo largo de un eje y también con precisión a lo largo de otro eje que está girado en algún ángulo con respecto al primero.

Eso solamente es trivial una vez que has aceptado la parte altamente no trivial: que la posición y el momento (o el espín en x y el

espín en z) no tienen una existencia independiente. Ambos, o, al menos, la probabilidad de observar que tengan un valor determinado, forman parte del estado cuántico único. En definitiva, es el *estado* lo que es físicamente relevante, no la forma en que decidamos expresarlo en una base u otra. Ese es el tema profundo.

Otra cosa que puede oír es que la mecánica cuántica se aplica al mundo de lo muy pequeño. En realidad, la mecánica cuántica se aplica a todo el mundo, grande y pequeño por igual. Pero existe una aproximación clásica que describe el mundo muy bien —extraordinariamente bien— cuando los objetos son suficientemente grandes. Esto no se deduce claramente de la ecuación (2.12), que no parece referirse en absoluto al tamaño del objeto. Pero sí se refiere al momento, que (para el movimiento no relativista, mucho más lento que la velocidad de la luz) se relaciona con la masa y la velocidad mediante $p = mv$. Generalmente, tomamos la masa como una cantidad fija y podemos pensar en la velocidad como un observable exactamente igual que el momento. Entonces podemos escribir la ecuación (2.12) como

$$\Delta x \, \Delta v \geq \frac{\hbar}{2m} . \tag{2.13}$$

Para un objeto masivo, la incertidumbre en los parámetros newtonianos x y v de toda la vida puede ser extremadamente pequeña cuando m es extremadamente grande. Esta es la razón por la que la mecánica clásica funciona tan bien en nuestro pesado mundo macroscópico, pero se viene abajo cuando consideramos partículas diminutas.

---

1 El término «espacio de Hilbert» fue introducido por el físico matemático húngaro John von Neumann y debe su nombre a David Hilbert, a quien conocimos en *Espacio, tiempo y movimiento* como colega de Einstein que dedujo la ecuación de campo de la relatividad general a partir del principio de mínima acción. La definición completa del espacio de Hilbert

implica otras condiciones matemáticas, incluida la capacidad de definir un «producto escalar» entre los vectores, pero no necesitaremos preocuparnos por esos detalles.

2 Sí, también hay «bras», representados con la notación $\langle\Psi|$. Podemos combinarlos para hacer un corchete de la forma $\langle\Psi|\Psi\rangle = 1$, que es la versión elegante en notación de Dirac de la ecuación (2.3). (N. de la T.: a diferencia de ket, bra tiene significado propio en inglés, y denota el «sostén» o «sujetador», en el sentido de la prenda interior femenina).

3 N. de la T.: El término «espín» procede del inglés *spin*, que en física significa «rotación sobre sí mismo».

TRES

ENTRELAZAMIENTO

El hecho de que las funciones de onda parezcan colapsarse cuando las observamos marca una profunda diferencia entre las formas clásica y cuántica de describir el mundo. Hace que a los físicos les resulte difícil ponerse de acuerdo sobre si la función de onda representa la realidad o es simplemente una herramienta para hacer predicciones. Pero esto no es lo único que hace que las funciones de onda parezcan de todo, menos intuitivas. Ni siquiera es lo más importante. Lo más importante es el hecho de que, en realidad, no hay *funciones* de onda, en plural, es decir, una función de onda para cada sistema físico individual. Únicamente hay una función de onda para todo en conjunto, lo que podríamos llamar (y llamamos) **la función de onda del universo**.

Eso es muy diferente de la mecánica clásica. Allí, cada parte del universo puede describirse por separado, y combinar sistemas es solamente cuestión de incluir la descripción adecuada para cada parte.

Supongamos que tenemos una partícula clásica, *A*, que se mueve en tres dimensiones. Todos los estados en los que podría encontrarse se recogen en un espacio de fases de seis dimensiones: tres dimensiones para la posición de *A* y tres dimensiones para su momento. Esa es la información que necesitamos para predecir su evolución futura. Supongamos ahora que añadimos otra partícula, *B*, al sistema. Esta partícula adicional está descrita por su propio espacio de fases de seis dimensiones: tres dimensiones para la posición de *B*, y lo mismo para su momento. El espacio de fases del sistema combinado es la suma de los dos espacios de fases, el de *A* y el de *B*: doce dimensiones en total. Si añadimos más subsistemas a nuestra colección, seguiremos agregando más dimensiones al espacio de fases. Cada partícula puede

interaccionar con las demás —lo cual es una cuestión dinámica que se aborda en las ecuaciones de movimiento—, pero la cuestión *cinemática* de cómo describir sus estados en un instante dado en el tiempo se responde simplemente con «un punto en un espacio de fases de seis dimensiones para cada partícula».

Siguiendo este patrón, podríamos imaginar que, si tenemos una partícula con una función de onda $\Psi(x)$, podríamos añadir otra partícula poniendo primero un subíndice en la función de onda de nuestra partícula original para convertirla en $\Psi_1(x)$, e introduciendo luego una nueva función de onda $\Psi_2(x)$ para la segunda partícula. No es así como funcionan las cosas en general. La coordenada x es un observable —un posible resultado de medición—, no un hecho sobre el sistema con algún valor definido. Así que deberíamos tener un observable x_1 para el lugar donde vemos la partícula A, y un observable separado x_2 para el lugar donde vemos la partícula B. Y la ecuación de Schrödinger $\widehat{H}\Psi = i\hbar^{\partial\Psi}/_{\partial t}$ solamente tiene cabida para una única función de onda, no para una combinación de dos o más. Por tanto, lo que realmente tenemos es una única función de onda que depende de las dos posiciones posibles en las que se pueden observar las partículas, $\Psi(x_1, x_2)$.

Y eso, a su vez, significa que lo que podamos observar sobre una partícula depende de lo que podamos observar sobre la otra, de una forma que en la mecánica clásica, simplemente, no ocurre. Este es el hecho con más implicaciones de la mecánica cuántica. Una vez que haya interiorizado esto realmente, muchas otras cosas se vuelven más claras.

La forma habitual en que los físicos expresan esta característica de la teoría cuántica es decir que dos partículas (u otros subsistemas de un sistema mayor) pueden estar **entrelazadas** entre sí. La terminología es imperfecta por un par de razones. En primer lugar, cuando el cerebro humano recibe la información de que dos cosas están «entrelazadas», no puede evitar visualizar alguna conexión tangible entre ellas, como una cuerda o un campo de fuerza o algo así. El

entrelazamiento es más sutil que eso. En concreto, las partículas no se ejercen fuerzas mutuamente por el mero hecho de estar entrelazadas. Y solamente midiendo una partícula concreta, no hay forma de saber si está entrelazada con otras.

En segundo lugar, y en un sentido más profundo, pensar en un sistema cuántico como un conjunto de partículas individuales que comparten una característica llamada «entrelazamiento» es una reliquia de la imagen clásica, en la que los estados de las partículas individuales tenían algún tipo de realidad independiente. No podemos evitar pensar en las partículas como lo que «realmente existe» y caracterizar su estado cuántico como entrelazado o no entrelazado. Sin embargo, es más útil pensar primero en el estado cuántico y elegir dividirlo en varias partículas como un paso posterior que resulta ser útil para los humanos, especialmente en el límite clásico.

Pero el término «entrelazamiento» está aquí para quedarse y seguiremos utilizándolo. La buena noticia —quizá— es que la confusión generada por el fenómeno del entrelazamiento y la confusión generada por la medición cuántica podrían anularse en última instancia, porque el entrelazamiento nos ayuda a entender la medición.

LA FUNCIÓN DE ONDA DEL UNIVERSO

Veamos por qué el entrelazamiento es una consecuencia natural de las características básicas de la mecánica cuántica.

La mejor manera de llegar a ello es pensar en la desintegración de una sola partícula en dos partículas. Pensemos en el bosón de Higgs, postulado en la década de 1960, pero no descubierto hasta 2012 en el experimento del Gran Colisionador de Hadrones. El bosón de Higgs es apasionante para los físicos de partículas por razones que explicaremos en capítulos posteriores, pero por ahora lo invocamos simplemente porque se desintegra para convertirse en otras partículas

de una manera sencilla: típicamente, en dos partículas a la vez, o, realmente, en una partícula y su antipartícula. Así que, a veces, se desintegrará en un electrón y su antipartícula el positrón; otras veces, en dos fotones (ya que los fotones son esencialmente sus propias antipartículas) y, a veces, en otras cosas.

Cuando hablamos de la desintegración del bosón de Higgs o de cualquier otra partícula inestable, tendemos a pensar que se trata de un proceso aleatorio con una cierta probabilidad de ocurrir en cualquier momento. Lo que realmente ocurre es que hay una función de onda cuántica que representa una superposición de estados de la forma «la partícula aún no se ha desintegrado» y «la partícula se ha desintegrado».

$$|\Psi(t)\rangle = \alpha(t)|\text{no desintegrada}\rangle + \beta(t)|\text{desintegrada}\rangle. \quad (3.1)$$

Cuando la observamos, esta función de onda colapsa y vemos la partícula única original o bien las dos partículas en las que se desintegra. Según la ecuación de Schrödinger, la amplitud $\alpha(t)$ disminuye con el tiempo y $\beta(t)$ aumenta. La probabilidad de que veamos la partícula original es $|\alpha|^2$, y la probabilidad de que veamos los dos productos de desintegración es $|\beta|^2 = 1-|\alpha|^2$. Así que, cuanto más esperemos, más probable será observar la partícula como desintegrada.

Centrémonos en el último término de la ecuación (3.1), el estado cuántico que representa las dos partículas en las que se ha desintegrado la original. Digamos que una de ellas es un electrón al que llamaremos 1, y la otra es un positrón al que llamaremos 2. Si realizamos el experimento muchas veces, observando las desintegraciones de muchos bosones de Higgs, veremos que los electrones y positrones creados se mueven en direcciones aleatorias alejándose de su punto de partida. Lo que esto significa —y lo que un examen cuidadoso de la ecuación de Schrödinger confirmaría— es que

la función de onda «desintegrada» se extiende en un patrón aproximadamente esférico alejándose de la partícula original.

Así que no podemos predecir en qué dirección se observará el electrón, ni en qué dirección se observará el positrón. Pero hay un dato más: el momento se conserva. Si nuestro bosón de Higgs original estaba simplemente ahí quieto sin momento, el momento del electrón observado debe ser igual en magnitud y opuesto en dirección al del positrón observado, de forma que los dos momentos sumen cero.

Pero espere un segundo. No podemos predecir la dirección en la que observaremos que se mueve el electrón, ni la del positrón. Supongamos que colocamos una pantalla detectora a un lado del bosón de Higgs original, y observamos que el electrón choca contra ella en algún punto. Comparando su posición con la localización de la partícula original, podemos averiguar la dirección en la que se movía el electrón emitido.

Y eso significa que ahora podemos deducir la dirección en la que se mueve el positrón, aunque no lo hayamos detectado. Debe moverse en dirección opuesta al electrón. De algún modo, la función de onda del positrón ha colapsado, pasando de ser una nube esférica que se mueve en todas direcciones a ser algo localizado con un momento determinado. ¿Cómo le ha ocurrido eso al positrón si ni siquiera lo hemos observado directamente?

La respuesta, como ya hemos revelado, es que el electrón y el positrón no se describen mediante estados cuánticos independientes. Existe un único estado para el sistema de dos partículas. Si x_1 es la localización donde observamos el electrón y x_2 es el positrón, por analogía con la ecuación (2.6) podemos escribir la función de onda combinada como $\Psi(x_1, x_2)$ y el estado en notación de Dirac como

$$|\Psi\rangle = \int \Psi(x_1, x_2) |\, x_1, x_2\rangle \, dx_1 \, dx_2, \qquad (3.2)$$

donde $|x_1, x_2\rangle$ es un estado base en el que cada partícula tiene una posición definida.

$\Psi(x_1, x_2)$ nos dice la probabilidad de que observemos simultáneamente el electrón en x_1 y el positrón en x_2, es decir, $P(x_1, x_2) = |\Psi(x_1, x_2)|^2$. Sin duda, podemos pensar en situaciones en las que dos partículas no tengan nada que ver la una con la otra, y su estado no esté entrelazado. Pero en otras situaciones, como cuando ambas partículas surgen de la desintegración de una sola, las probabilidades están muy relacionadas. Eso es el entrelazamiento.

Este patrón es generalizado. No importa lo grande que sea nuestro sistema, solamente hay una función de onda. En última instancia, solamente hay un estado cuántico para todo, la función de onda del universo. A menudo es posible subdividirla porque resulta que diferentes trozos no están entrelazados, pero esa no es la situación genérica.

Hay que destacar dos cosas sobre esta historia del entrelazamiento. En primer lugar, deja claro que las funciones de onda no son campos, en el sentido en que hemos hablado de campos en la mecánica clásica. La idea principal de un campo es que toma un valor único en cada punto del espacio. Para una sola partícula, cuando la función de onda se parece a $\Psi(x)$, eso es básicamente lo que ocurre, por lo que se le perdonaría que pensara en la función de onda como una especie de campo. Pero en cuanto tenemos dos partículas, $\Psi(x_1, x_2)$ no asigna un número a cada punto del espacio: asigna un número a cada *par* de puntos. Y para tres partículas, la función de onda $\Psi(x_1, x_2, x_3)$ asigna un número a cada triplete de puntos, y así sucesivamente. Las funciones de onda no son funciones del espacio, sino que son funciones del *espacio de configuración* del sistema que estemos considerando. (O del espacio de momentos, si queremos representarlas así). Eso significa que las funciones de onda no son campos en el sentido habitual. Esto es lo que permite la posibilidad de entrelazamiento.

Otra cosa que hay que tener en cuenta es que la ecuación (3.1), que representa la superposición de una sola partícula sin desintegrar y sus dos productos de desintegración, huele un poco a chamusquina si la

estudiamos detenidamente. El espacio de Hilbert para una sola partícula está formado por funciones de valor complejo de una variable, $\Psi(x)$, mientras que el espacio de Hilbert para dos partículas está formado por funciones de valor complejo de dos variables, $\Psi(x_1, x_2)$. ¿Es legítimo juntarlas sencillamente en un único espacio de Hilbert?

La respuesta es que se puede legitimar, pero para describir las transiciones entre distintos números de partículas tenemos que recurrir a la teoría cuántica de campos. Pronto llegaremos allí.

ACCIÓN FANTASMAL A DISTANCIA

El entrelazamiento se destacó por primera vez como una característica de la mecánica cuántica en un famoso artículo de 1935 de Albert Einstein, Boris Podolsky y Nathan Rosen, que posteriormente se conoció simplemente como EPR. Surgió de la convicción de Einstein de que la mecánica cuántica no podía ser la historia final correcta en lo que respecta al funcionamiento del universo. El artículo incidía en que no se trataba de que la mecánica cuántica fuera errónea, sino de que estaba incompleta. Lo dice claramente en el título: «¿Puede considerarse completa [la] descripción cuántico-mecánica de la realidad física?». (La ausencia de la palabra «la»[1] refleja el estilo propio de la revista *Physical Review*, no el hecho de que Podolsky fuera ruso. Creo).

El EPR consideró una situación muy parecida a la de nuestra partícula en desintegración, pero es más fácil explicar el punto básico utilizando espines entrelazados en lugar de posiciones de partículas. Un bosón de Higgs tiene espín total cero, mientras que el electrón y el positrón son partículas de espín ½. Por conservación del momento angular, el electrón y el positrón deben tener espines opuestos, ya que deben sumar cero.

El entrelazamiento nos permite considerar un estado en el que el espín de cualquiera de las dos partículas es completamente incierto, pero los dos están obligados a ser opuestos. Consideremos la medición del espín en la dirección del eje z, y escribamos un estado de espín hacia arriba como $|\uparrow\rangle$ y un estado de espín hacia abajo como $|\downarrow\rangle$. Para el sistema de dos partículas electrón/positrón, podemos definir para ambas cuatro estados base de espín determinado, donde la primera flecha representa al electrón y la segunda al positrón:

$$|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle .$$

Un estado como $|\uparrow\downarrow\rangle$ significa que tenemos garantizado encontrar el espín hacia arriba del electrón y el espín hacia abajo del positrón cuando los medimos a lo largo del eje z. Tiene momento angular total cero, pero los dos espines *no* están entrelazados: en ese estado no aprenderíamos nada nuevo sobre el electrón midiendo el positrón, o viceversa. Sin embargo, al tomar la superposición correcta, podemos construir un estado en el que ninguno de los espines se conoce de antemano, pero se garantiza que el momento angular total es cero:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{3.3}$$

Este es el tipo de estado en el que se encontrarán los espines cuando las partículas sean emitidas por un bosón de Higgs en desintegración. Si medimos el espín del electrón, la probabilidad de que veamos el estado $|\uparrow\downarrow\rangle$ viene dada por su amplitud al cuadrado, $\left(1/\sqrt{2}\right)^2 = 1/2.$ En ese caso, el electrón tiene espín hacia arriba. El estado en su conjunto habrá colapsado para pasar a la parte en la que el electrón está con espín hacia arriba, es decir, $|\uparrow\downarrow\rangle$. Así que ahora sabemos con certeza que el positrón tendría espín hacia abajo si también lo medimos en la dirección del eje z. Del mismo modo, si vemos que el electrón tiene espín hacia abajo, sabemos que el positrón

tendrá espín hacia arriba. La medición de una parte de la función de onda ha arrastrado al resto, debido al entrelazamiento.

Existe un mito popular según el cual Einstein, por muy gran físico que fuera, se había convertido en una persona chapada a la antigua cuando floreció la mecánica cuántica, y diversos prejuicios filosóficos le impidieron apreciar plenamente la teoría. Nada más lejos de la realidad. Como hemos visto, Einstein ayudó a lanzar la revolución cuántica, y en 1935 entendía la mecánica cuántica tan bien como cualquiera. Simplemente no creía que fuera la respuesta definitiva. Y, aunque es cierto que una vez afirmó, a modo de queja, que Dios no juega a los dados con el universo, la naturaleza aleatoria de las predicciones cuánticas no era su principal escollo. Le interesaba mucho más el *realismo* —el hecho de que realmente hay un mundo físico ahí fuera, independiente de nuestras observaciones— y la *localidad* —el hecho de que existe algo llamado «espacio-tiempo», y lo que ocurre en el mundo ocurre en localizaciones individuales del espacio-tiempo.

El EPR pensó que la experimentación ponía al realismo y la localidad en conflicto, lo cual implica lo que Einstein bautizó memorablemente como **acción fantasmal a distancia**. Tras la desintegración del bosón de Higgs, el electrón y el positrón se alejan entre sí. Podríamos capturarlos sin medir sus espines, mantener uno cerca y enviar al otro en un cohete espacial que viajara a años luz de distancia. Digamos que Alicia está aquí en un laboratorio en la Tierra con el electrón, y Roberto está orbitando Alfa Centauri con el positrón. Alicia finalmente mide el espín del electrón, e imaginemos que lo registra como espín hacia arriba. Antes de esa medición, no teníamos ni idea de lo que Roberto vería para el positrón. Después, Alicia sabe con certeza que él observará un espín hacia abajo. El estado del positrón ha cambiado instantáneamente, no importa lo lejos que estuviera. Einstein, de entre toda la gente, sabía perfectamente que ninguna información real puede viajar entre las partículas más rápido que la velocidad de la luz, por las reglas de la

relatividad. La palabra «instantáneamente» ni siquiera está bien definida en lugares muy separados. Entonces ¿cómo «sabe» el positrón cómo responder a la medición del electrón lejano?

Esto resulta más bien ir en contra del espíritu de la relatividad que ser una verdadera ley escrita. Por supuesto, Alicia sabe también cuál será el espín del positrón en cuanto mide el del electrón. Pero Roberto no se entera de nada que no supiera ya, puesto que está demasiado lejos para ver lo que Alicia ha medido. Por lo que a él respecta, todavía hay una probabilidad de ½ de ver espín hacia arriba o espín hacia abajo cuando mida el positrón. Alicia podría enviarle un mensaje diciéndole cuál fue el resultado de su medición, pero ese mensaje tendría que viajar por medios ordinarios, más lentos que la luz. A pesar de la imaginación activa de la gente que sabe un poco de mecánica cuántica, pero no lo suficiente, el entrelazamiento no nos permite enviar información a una velocidad superior a la de la luz.

EL PUZLE DEL EPR

Existe la tentación de imaginar que tal vez el entrelazamiento no sea tan fantasmal como lo pintan. Imagine que tiene dos canicas, una roja y otra azul, y que coloca cada una de ellas en cajas de aspecto idéntico. Un amigo mezcla las dos cajas para que no sepa cuál es cuál, le da una y se lleva la otra a un lugar lejano. Usted abre su caja y ve la canica roja. Ahora sabe al instante que la otra es azul, aunque esté lejos y no lo supiera antes. ¿Qué tiene eso de fantasmal?

Esto no es lo que ocurre en la mecánica cuántica. El entrelazamiento es algo más que una simple correlación entre cosas que ya existen, pero que desconocemos. Lo sabemos en parte porque la función de onda actúa como algo físicamente real, por ejemplo, al presentar interferencias en el experimento de la doble rendija. Pero, más directamente, el tipo de correlaciones que implica el

entrelazamiento cuántico va mucho más allá de las simples relaciones clásicas.

El EPR se dio cuenta de ello, así que intentaron convertir su incomodidad en algo más explícitamente paradójico. Postulan que, cuando podemos decir con absoluta certeza cuál será el resultado de una medición, debe haber un «elemento de realidad» que corresponda a esa propiedad del sistema. Según esta definición, después de que Alicia mida que el electrón tiene espín hacia arriba, hay un elemento de la realidad que dice «el positrón de Roberto tendrá espín hacia abajo». Y según el EPR, ese elemento debe haber estado ahí todo el tiempo, ya que nada puede viajar más rápido que la luz.

Pero Alicia no tenía que medir el espín de su electrón en la dirección del eje z; también podría haberlo medido en la dirección del eje x, obteniendo posibles resultados «espín hacia la derecha» o «espín hacia la izquierda» en lugar de hacia arriba y hacia abajo. Vamos a etiquetar los estados correspondientes $|\rightarrow\rangle$ y $|\leftarrow\rangle$ (que llamamos $|+x\rangle$ y $|-x\rangle$ en el capítulo anterior). Comparando la relación entre los espines en la dirección del eje z y los espines en la dirección del eje x de (2.7) con el estado entrelazado (3.3), es bastante fácil verificar que el estado también se puede escribir como

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|\leftarrow\rightarrow\rangle - |\rightarrow\leftarrow\rangle\right). \tag{3.4}$$

Este es el mismo estado cuántico que (3.3), solamente que escrito en una base diferente. Los dos espines siguen entrelazados y siguen teniendo alineaciones opuestas: si Alicia mide espín hacia la izquierda, Roberto verá inevitablemente espín hacia la derecha, y viceversa. Eso tiene sentido, dado que se supone que el momento angular debe sumar cero.

El EPR afirmaba que es fácil imaginar una situación en la que Alicia habría obtenido espín hacia arriba si hubiera medido el espín en la dirección del eje z, y habría obtenido espín hacia la izquierda si

hubiera medido el espín en la dirección del eje x. Por lo tanto, en opinión de los autores, debe haber elementos separados de la realidad que fijen tanto el espín en z como el espín en x de la partícula de Roberto, y (debido a la relatividad) esos elementos deben haber estado ahí todo el tiempo. Pero eso se opone al principio de incertidumbre. De acuerdo con las reglas tradicionales de la mecánica cuántica, no se puede tener un estado que se encuentre al mismo tiempo en un valor definido de espín en z y de espín en x. Por lo tanto, concluyó el EPR, la propia mecánica cuántica debe estar incompleta.

Frente al argumento del EPR, la respuesta estándar de un físico moderno es simplemente aceptar que existe algún tipo de influencia que se origina en el evento de medición de Alicia y se propaga instantáneamente (signifique eso lo que signifique) al de Roberto, aunque los dos eventos estén separados a distancias a escalas espaciales (fuera de los conos de luz de cada uno). Como se ha mencionado, no hay forma de aprovechar esta fantasmal acción a distancia para enviar señales o información, aunque el formalismo que utilicemos parezca invocar influencias más rápidas que la luz. Parte de la respuesta estándar consiste en no hacer demasiadas preguntas incómodas sobre esta situación y cambiar de tema.

MEDICIÓN Y ENTRELAZAMIENTO

Al principio de este capítulo hemos insinuado la posibilidad de que el entrelazamiento nos ayudara a comprender el misterio de la medición cuántica. Veamos cómo pueden estar relacionados ambos.

Para conseguirlo, seamos más explícitos sobre lo que ocurre durante una medición. Hemos dicho que un sistema cuántico puede estar en una superposición, pero cuando se mide una cantidad, solamente vemos que esa cantidad tiene un valor definido. Para realizar este tipo de medición necesitamos algún tipo de aparato. En el experimento de Stern-Gerlach que mide el espín, por ejemplo,

introducimos una partícula en una superposición de espín hacia arriba y espín hacia abajo, y el aparato da una respuesta definitiva, «hacia arriba» o bien «hacia abajo». Podemos ser aún más concretos e imaginar un dial con una aguja grande y las etiquetas *ARRIBA*, *Listo* y *ABAJO*. Antes de la medición, la aguja apunta a *Listo*, lo que indica que aún no hemos empezado. Después de la medición, apunta hacia *ARRIBA* o bien hacia *ABAJO*, dependiendo del resultado.

Es de suponer que nuestro aparato está hecho de átomos, partículas, y similares, por lo que debería obedecer las reglas de la mecánica cuántica igual que lo haría una única partícula giratoria. La descripción cuántica completa de un aparato macroscópico de laboratorio puede ser intimidatoriamente complicada, pero no perderemos nada importante para nuestros propósitos actuales reduciéndola a tan solo tres estados básicos: $|ARRIBA\rangle$, $|Listo\rangle$ y $|ABAJO\rangle$, cada uno asociado a la correspondiente posición definitiva del puntero. Por supuesto, también tenemos libertad para imaginar superposiciones de estos estados, pero nunca nos encontramos con ellas en laboratorios reales. Los punteros con los que estamos familiarizados siempre parecen apuntar en direcciones definidas.

Podemos empezar con una partícula en alguna superposición, $|\Psi\rangle_{\text{partícula}} = \alpha|\uparrow\rangle+\beta|\downarrow\rangle$, y luego dejar que sea medida por un aparato en el estado inicial $|\Psi\rangle_{\text{aparato}} = |Listo\rangle$. El estado inicial completo es simplemente el producto de estos dos, que podemos escribir como

$$|\Psi\rangle_{\text{inicial}} = (\alpha|\uparrow\rangle+\beta|\downarrow\rangle)\,|Listo\rangle. \qquad (3.5)$$

Podríamos considerar situaciones iniciales en las que la partícula y el aparato ya estuvieran entrelazados, pero vamos a optar por considerar una sin entrelazamiento.

Cuando el proceso de medición se ha completado, el asunto ha colapsado para tener la partícula en un estado completamente $|\uparrow\rangle$ y el aparato en el estado $|ARRIBA\rangle$, o la partícula en un estado completamente $|\uparrow\rangle$ y el aparato en el estado $|ABAJO\rangle$.

$$|\Psi\rangle_{\text{final}} = |\uparrow\rangle|ARRIBA\rangle, \quad \textit{probabilidad } |\alpha|^2,$$

$$|\Psi\rangle_{\text{final}} = |\downarrow\rangle|ABAJO\rangle, \quad \textit{probabilidad } |\beta|^2. \qquad (3.6)$$

Todo este asunto del colapso, como sabemos, es una vulneración flagrante de la ecuación de Schrödinger, y no está claro cómo, por qué o precisamente cuándo se produce el colapso.

Así que, preguntémonos: ¿qué pasaría si *no* vulnerásemos la ecuación de Schrödinger? Se podría suponer que es una pregunta tan difícil que es imposible responderla, ya que el sistema de átomos y moléculas que componen un aparato macroscópico es demasiado complejo para caracterizarlo con precisión. Pero tenemos ante nosotros una enorme simplificación, a partir del hecho básico de que queremos que nuestra medición sea precisa. Esto significa que, si el estado de la partícula inicial era $|\Psi\rangle_{\text{particle}} = |\uparrow\rangle$ (espín hacia arriba puro, sin combinación con espín hacia abajo), entonces el estado final del aparato tiene que ser $|ARRIBA\rangle$ puro, con probabilidad del 100%, y lo mismo para el espín hacia abajo. La única manera de que esto ocurra es que el estado inicial más general (3.5) evolucione a

$$|\Psi\rangle_{\text{final}} = \alpha|\uparrow\rangle|ARRIBA\rangle + \beta|\downarrow\rangle|ABAJO\rangle. \qquad (3.7)$$

En otras palabras, lo que le ocurre a un buen dispositivo de medición es que se entrelaza con el sistema que se está midiendo, de tal forma que asocia las lecturas adecuadas con los estados correspondientes del sistema.

Sin embargo, eso no es lo que ven realmente los científicos del mundo real. Lo que vemos son indiscutiblemente partículas y punteros en un estado definido, como en la ecuación (3.6). Por tanto, el problema de la medición no es que se resuelva, sino que se define mejor: ¿qué perturba la evolución ordinaria de Schrödinger para que aparentemente acabemos con (3.6) en lugar de (3.7)?

DECOHERENCIA

Hay un cabo suelto. Hemos insistido mucho en el hecho de que solamente hay una función de onda para todo el universo, pero hemos analizado continuamente situaciones que contienen solo unas pocas partes móviles, ignorando todo lo demás que hay en el mundo. Eso es aceptable cuando pensamos en partículas microscópicas, que pueden separarse del resto del mundo durante un tiempo razonable. Pero cuando se trata de un objeto macroscópico, como un aparato de medición, esta maniobra de la «vaca esférica» se rompe. Nos guste o no, partes del resto del mundo —moléculas en el aire, fotones ambientales en la habitación— van a seguir chocando contra el aparato. ¿Importa eso?

Pues sí. Volvamos a analizar la medición de un solo espín, pero ahora incluyamos todo lo demás que hay en el universo, envuelto en un único sistema que denominamos **entorno**. Eso incluye moléculas de aire, fotones, pero también al experimentador, la mesa y todo lo demás. Considere que el entorno son todas esas cosas que no controlamos explícitamente durante el experimento. Una vez más, el estado del entorno será inimaginablemente complejo, pero solo nos interesa un pequeño número de características.

Piense en cómo interacciona nuestro aparato de medición con su entorno. El experimentador, si tiene cuidado, casi siempre dejará solo el aparato durante la medición, por lo que no ocurrirá nada interesante. Pero los fotones de la habitación serán importantes, ya que los necesitamos para iluminar el dial. Y —esta es la parte crucial — esos fotones interaccionarán de forma diferente con el aparato dependiendo de lo que este esté haciendo. Hemos imaginado un puntero que podría apuntar a las diferentes lecturas: *ARRIBA, Listo* y *ABAJO*. Digamos que el puntero es negro (por lo que absorbe fotones), mientras que el fondo del dial es blanco (por lo que los refleja). Entonces es fácil imaginar cómo un determinado fotón chocará el puntero si está apuntando hacia *ARRIBA* y será absorbido,

pero el mismo fotón será reflejado por el dial si el puntero apunta a alguna dirección diferente.

Esto, a su vez, significa que el baño de fotones en el que necesariamente está sumergido el aparato no solo interaccionará con este, sino que se entrelazará con él, casi inmediatamente.[2] Digamos que el entorno comienza en un estado inicial $|e_0\rangle$, y luego evoluciona a $|e_\uparrow\rangle$, o bien a $|e_\downarrow\rangle$, dependiendo de si sus fotones chocan con una flecha que apunta hacia *ARRIBA* o hacia *ABAJO*, respectivamente. Entonces, el estado combinado de espín/aparato/entorno inmediatamente después de la medición tendrá un aspecto similar a

$$|\Psi\rangle_{\text{medición}} = \left(\alpha|\uparrow\rangle\,|ARRIBA\rangle + \beta|\downarrow\rangle\,|ABAJO\rangle\right)|e_0\rangle, \quad (3.8)$$

pero los fotones se mueven a la velocidad de la luz, así que esto evolucionará extremadamente rápido para convertirse en

$$|\Psi\rangle_{\text{decoherente}} = \alpha|\uparrow\rangle\,|ARRIBA\rangle\,|e_\uparrow\rangle + \beta|\downarrow\rangle\,|ABAJO\rangle\,|e_\downarrow\rangle. \quad (3.9)$$

Por tanto, no solamente el aparato de medición se entrelaza con el sistema que se está midiendo. Como el aparato es un objeto macroscópico que no puede evitar interaccionar con el baño de fotones que lo rodea, el entorno también se entrelaza. Esto se conoce como **decoherencia**: proceso por el cual un sistema cuántico en una superposición se entrelaza con su entorno. Para objetos grandes que continúan chocando con el resto del mundo, la decoherencia es rápida y casi inevitable.

¿Por qué debería importarnos que los sistemas macroscópicos se entrelacen continuamente con su entorno? Recordemos el experimento de la doble rendija. Cuando no observábamos la partícula al pasar por las rendijas, veíamos cómo aparecían patrones de interferencia en la pantalla del detector del otro lado. Esta interferencia se debe a que hay dos contribuciones —una para cada rendija— precisamente a la misma amplitud final para una partícula que se observa en una

posición determinada. Cuando la partícula se «mide» mientras atraviesa las rendijas, en realidad se está entrelazando con algún dispositivo de medición externo. Esto significa que, cuando llega a la pantalla del detector, las contribuciones de las dos rendijas no son directamente comparables. Están entrelazadas con diferentes estados del mundo exterior, por lo que no pueden ni anularse mutuamente ni sumarse. La decoherencia destruye la interferencia, que es uno de los aspectos más fundamentales del comportamiento cuántico. Y la decoherencia es irreversible: el entorno tiene un conjunto tan grande y caótico de partes en movimiento que, una vez que se entrelaza con un sistema, la posibilidad de que se desentrelace en algún momento es insignificante.

Esto sugiere una respuesta a lo que una medición cuántica es realmente: el proceso por el cual el sistema se descohesiona. La medición, tal y como aparece en la mecánica cuántica, no tiene nada que ver con la conciencia o la percepción. Es simplemente lo que ocurre cuando un sistema cuántico en una superposición se entrelaza con su entorno.

FUNDAMENTOS

Eso no significa que la decoherencia «resuelva el problema de la medición». Explica por qué desaparece el patrón de interferencia en el experimento de la doble rendija y, por tanto, sugiere una buena comprensión basada en la física de lo que queremos decir con «medición». Pero, a fin de cuentas, seguir simplemente la ecuación de Schrödinger nos deja con el estado entrelazado (3.9), mientras que lo que aparentemente vemos es uno de los estados colapsados de (3.6). Todavía tenemos que explicar de alguna manera por qué vemos esto último en lugar de lo primero.

Este es el tema de los **fundamentos de la mecánica cuántica**: ¿cuál es la teoría completa y sin ambigüedades que explica las predicciones

correctas del formalismo un tanto *ad hoc* de «mediciones» vagamente definidas y «colapsos» repentinos? Aún no estamos de acuerdo en cuál es la respuesta. Aquí, en *Las ideas fundamentales del universo*, nos centramos en las cosas en las que sí estamos de acuerdo, así que nos limitaremos a ofrecer una rápida visión general de algunas de las ideas más destacadas.

He aquí una: tal vez la función de onda represente la realidad y obedezca siempre a la ecuación de Schrödinger. Eso estaría muy bien, desde el punto de vista de la estética y la simplicidad. El problema es que acabamos obteniendo la ecuación (3.9): un estado cuántico que está en una superposición de diferentes resultados de medición. Y ningún experimentador ha tenido nunca la sensación de encontrarse en una superposición de diferentes resultados de medición. Los experimentos reales parecen tener resultados definidos.

Una forma de resolver este problema fue sugerida por Hugh Everett III cuando era estudiante de posgrado en la década de 1950. Everett propuso que la ecuación de Schrödinger está bien, lo que está mal es nuestra identificación del «experimentador» en un estado como (3.9).

Recordemos que, una vez que se produce la decoherencia, en la práctica las partes del estado que están entrelazadas con diferentes estados del entorno nunca volverán a interferir entre sí. En ese caso, dice Everett, lo correcto es tratarlas como copias completamente independientes del universo, llamadas **ramas**. No hay ningún experimentador que esté en una superposición; hay un experimentador en la rama $|ARRIBA\rangle$ de la función de onda, y otro en la rama $|ABAJO\rangle$, y ambos experimentan resultados de medición definidos. Esto se conoce como la **interpretación de los Muchos Mundos** de la mecánica cuántica. (Aunque es una teoría física formal por derecho propio, no una «interpretación» de algo).

La interpretación de los Muchos Mundos es sencilla y convincente, pero levanta ampollas por razones obvias. Un enfoque diferente consiste en plantear que la función de onda representa una *parte* de la

realidad, pero no toda: que hay otras variables necesarias para describir completamente la configuración física del mundo. En el caso de una colección de partículas, esas variables podrían ser simplemente las posiciones reales de las partículas. Este enfoque fue promovido por el propio Louis de Broglie, y más tarde recuperado y popularizado por David Bohm, por lo que a menudo se conoce como **teoría de De Broglie-Bohm** o, simplemente, **mecánica bohmiana**. Es una especie de **teoría de variables ocultas**, porque necesitamos variables adicionales que nos digan dónde están las partículas. También se conoce como teoría **de ondas piloto**, porque el papel de la función de onda es «pilotar» las partículas. Estas tienden a congregarse allí donde la función de onda es grande y a dispersarse del lugar donde es pequeña.

Las teorías de variables ocultas tienen una historia curiosa. En 1935, John von Neumann presentó una supuesta prueba de que nunca podrían funcionar. Grete Hermann señaló que von Neumann había hecho suposiciones innecesarias, pero fue ignorada en gran medida y los físicos asumieron que la posibilidad había sido eliminada. En la década de 1950, Einstein apuntó las limitaciones del argumento de von Neumann a Bohm, que se puso a inventar un modelo viable y llegó a uno similar al de De Broglie de años antes. Este también fue ignorado en gran medida, pero llamó la atención del físico irlandés John Bell, teórico del CERN, el laboratorio de física de partículas de Ginebra (Suiza). Bell observó una característica importante de la teoría de Bohm: era **no local**, porque la forma en que la función de onda empujaba las partículas dependía de la configuración de todas las partículas, no solamente de la que era empujada. Bell se preguntó si esta característica era necesaria y demostró que lo era: el **teorema de Bell** establece que es imposible reproducir las predicciones de la mecánica cuántica en una teoría de variables ocultas que sea completamente local.[3] El Premio Nobel de Física fue concedido en 2022 a Alain Aspect, John Clauser y Anton Zeilinger por demostrar experimentalmente que las partículas entrelazadas en el tipo de configuración que Bell imaginó realmente obedecían a las predicciones

de la teoría cuántica estándar. Por desgracia, en el comunicado de prensa del Nobel se cometió un error al afirmar que el teorema de Bell eliminaba la posibilidad de modelos de variables ocultas. En realidad, solamente elimina la posibilidad de modelos *locales* de variables ocultas. El propio Bell se habría sentido atormentado, ya que era un gran admirador de la teoría no local de Bohm.

Otro enfoque viable de la medición cuántica es simplemente aceptar que las funciones de onda en verdad colapsan, incluso si esto vulnera la ecuación de Schrödinger. Existen varias estrategias en este sentido, dependiendo de si los colapsos son espontáneos y aleatorios o están provocados por alguna característica del estado cuántico. La buena noticia es que estos **modelos de colapso objetivo** se pueden probar experimentalmente. La mala noticia, al menos hasta ahora, es que no hay pruebas experimentales a su favor.

Por último, algunos físicos prefieren volver a la filosofía original de Copenhague de Bohr y Heisenberg. La idea es dejar de preocuparse por lo que es «real» y centrarse en lo que realmente vemos: los resultados de las mediciones. En los **enfoques epistemológicos** de los fundamentos cuánticos, las funciones de onda son meras herramientas para que los interesados calculen las probabilidades de los resultados de las mediciones. Por tanto, su colapso no tiene nada de fantasmal; es simplemente el proceso habitual de actualizar nuestras creencias cuando aparecen nuevas pruebas.

La variedad de opciones sobre la mesa refleja la incapacidad de la comunidad de físicos para ponerse de acuerdo sobre los fundamentos de la mecánica cuántica. A los físicos se les da extremadamente bien utilizar una teoría para hacer predicciones, pero no tanto establecer lo que realmente dice la teoría.

Para nuestros propósitos en este libro, tenemos por tanto una situación: ¿cómo vamos a hablar de las funciones de onda y de la realidad, dado que los físicos profesionales no se ponen de acuerdo sobre la forma correcta de hacerlo? Podríamos, en cada coyuntura relevante, indicar una serie de maneras diferentes de expresar lo que

está sucediendo fundamentalmente, dependiendo de la actitud de cada uno hacia los fundamentos de la mecánica cuántica. Pero eso sería tedioso. En su lugar, vamos a elegir una actitud y tratar de ser coherentes al respecto: a saber, que la función de onda representa directa e inequívocamente la realidad. Y, como veremos, esa función de onda se construye a partir de campos cuánticos que se extienden por todo el espacio-tiempo. En esta imagen, los electrones no son realmente partículas puntuales, sino vibraciones cuantizadas en el campo electrónico; los átomos no son en su mayor parte espacio vacío, sino que están definidos por un determinado perfil de función de onda. Cuando lleguemos a las interacciones y a los diagramas de Feynman, mencionaremos que las «partículas virtuales» son aún menos reales literalmente; son un instrumento cómodo para calcular la probabilidad de ciertos procesos cuánticos. Pero son en verdad un instrumento muy cómodo, y es del todo correcto hablar de ellas, siempre y cuando recordemos lo que realmente está pasando al observar bajo la lupa.

Tenga en cuenta que existen perspectivas alternativas. Le invitamos a traducir el lenguaje utilizado aquí a cualquier cosa que sugiera su enfoque favorito de los fundamentos cuánticos. No cabe duda de que las generaciones futuras podrán aclarar los aspectos importantes de la situación.

---

1 N. de la T.: En inglés, falta el artículo «the» en el título: «Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?».

2 Se puede interaccionar sin entrelazarse; el entrelazamiento solamente se produce cuando diferentes estados de un sistema se asocian con diferentes estados de otro. El estado (3.7) está entrelazado, mientras que (3.5) no lo está.

3 Como cualquier teorema matemático, el de Bell viene con suposiciones. Una de ellas es que los experimentos tienen resultados definidos, lo que se incumple en la interpretación de los Muchos Mundos, donde se producen diferentes resultados en diferentes ramas. Otra es que podemos medir cualquier cantidad que queramos medir; esto se incumple en las teorías «superdeterministas», en las que las condiciones iniciales del universo garantizan que ciertas propiedades nunca se medirán.

CUATRO

CAMPOS

Al igual que la mecánica clásica, la mecánica cuántica no es una teoría física específica en sí misma. Es un marco dentro del cual podemos construir modelos específicos de distintos tipos de sistemas. Existe una teoría clásica del oscilador armónico simple y existe también una teoría cuántica del oscilador armónico simple. Muy a menudo, en física tenemos en mente un tipo de sistema físico y luego pensamos en cómo construir una teoría cuántica de ese sistema.

En mecánica clásica distinguimos entre partículas, que tienen posiciones concretas en el espacio, y campos, que toman valores en cada punto del espacio. Podemos hacer lo mismo en mecánica cuántica. En los últimos capítulos hemos estudiado las teorías cuánticas de las partículas. La teoría mecánica cuántica de los campos se conoce acertadamente como **teoría cuántica de campos** o **TCC**. Algunas personas hablan como si primero hubiera existido la «mecánica cuántica» y luego hubiera sido sustituida por la «teoría cuántica de campos». Eso no es cierto: la TCC es parte de la mecánica cuántica, solo que aplicada a campos en lugar de a partículas.

En el momento de escribir estas líneas, la teoría cuántica de campos es la mejor forma que tenemos de describir el universo en su nivel más profundo conocido. Ni siquiera es necesario que consideremos las partículas y los campos por separado; cuando pensemos detenidamente en los campos cuánticos, las partículas aparecerán frente a nosotros sin más. Es el ejemplo por excelencia de la aparición de «cuantos» que surgen a partir de una imagen del mundo fundamentalmente continua.

Consideremos un campo simple, que denotaremos por ϕ (letra griega phi). Al ser un campo, tendrá un **valor** en cada punto x del espacio, $\phi(x)$. Ese valor puede ser cero, pero sigue siendo un valor. (Por razones de simplicidad en la notación, vamos a escribir todo como si solamente hubiera una dimensión del espacio, x, pero todas las ideas funcionan igual de bien en tres dimensiones). Incluso si el espacio está lo más vacío posible, los campos siguen estando ahí. Esto contrasta con las partículas, que tienen una ubicación en el espacio en cada instante de tiempo, pero ningún tipo de existencia en los demás puntos en los que no se ubican.

A menudo resulta útil considerar una **configuración de campo** completa, o «perfil», que especifica los valores reales que tiene el campo en cada punto del espacio. Denotaremos las configuraciones de campo mediante corchetes, $\{\phi(x)\}$. Pero a veces somos perezosos y nos saltamos los corchetes (y otras referencias no los utilizan en absoluto), así que hay que prestar atención. A medida que el campo evoluciona en el tiempo de acuerdo con una ecuación de movimiento adecuada, alcanzará valores $\phi(x,t)$ en cada punto del espacio-tiempo. Matemáticamente, podríamos decir que un campo está definido por un mapeo entre el espacio y algún tipo de valores. De momento nos ceñiremos al caso más sencillo, un **campo escalar**, en el que los valores son simplemente números en cada punto del espacio. Más adelante estudiaremos los campos vectoriales y otras posibilidades más exóticas.

Como veremos en breve, en realidad, la mejor manera de llegar a entender los campos cuánticos no es pensar en ellos como valores en cada punto del espacio, sino expresar una configuración de campo como una suma de ondas con diferentes longitudes de onda (mediante la transformada de Fourier). Cada término de esa suma es un «modo» de longitud de onda definida, muy parecido a una nota musical de tono definido.

Y entonces se produce el milagro. Cada modo de un campo cuántico se comporta como un oscilador armónico simple, incluyendo

los niveles de energía cuantizada que descubrimos previamente. Esos niveles de energía se interpretan como el número de *partículas* que observaríamos: un modo en su primer estado excitado representa una partícula, su segundo estado excitado representa dos partículas, y así sucesivamente. Se trata de una conexión cuántica bastante mágica entre campos y partículas, y merecerá la pena ensuciarse un poco las manos en la maleza matemática para ver cómo se produce.

Es tentador preguntarse qué «son realmente» los campos. Por ejemplo, ¿de qué están hechos? No hay una respuesta satisfactoria a esta pregunta. En el contexto de la teoría cuántica de campos, los campos no están «hechos de» nada: son las cosas de las que está hecho todo lo demás. Esto no debería sorprendernos. Si indagamos lo suficiente sobre de qué están hechas las cosas, llegaremos a la conclusión de que «aquí está la materia esencial de la realidad, que no está hecha de nada más».

Según nuestros mejores conocimientos actuales, los campos cuánticos son la materia esencial de la realidad. Es completamente concebible que en el futuro nuestra comprensión mejore y lleguemos a apreciar que los campos no son fundamentales después de todo. Incluso hay indicios de ello en los intentos modernos de comprender la gravedad cuántica. Pero eso sigue siendo bastante especulativo, y para los fines de las Ideas Fundamentales tiene sentido ceñirse al paradigma establecido de los campos cuánticos.

LA ENERGÍA DE UN CAMPO

La mayoría de las veces, cuando los físicos construyen una teoría cuántica, parten de una teoría clásica y la cuantizan. Esto es cierto tanto para los campos como para las partículas. Así que pensemos en un campo clásico y preguntémonos qué tipo de energía podría llevar dicho campo. Eso nos llevará a un hamiltoniano cuántico apropiado, que da sentido a la ecuación de Schrödinger.

Para una sola partícula, previamente consideramos dos tipos de energías: la energía cinética $\frac{1}{2} mv^2$, y la energía potencial $V(x)$, donde m es la masa de la partícula, v es su velocidad y $V(x)$ es la función de energía potencial. La energía cinética está asociada al movimiento, y la energía potencial solamente depende de la posición de la partícula y de los factores externos que influyan en ella. Cuando se trata de campos, no hay nada parecido a «la posición», sino que existe el valor $\phi(x,t)$ que el campo tiene en cada evento (x,t), así como los valores en otros puntos del espacio y del tiempo. Necesitamos construir algo parecido a la energía a partir de esa información.

Una de las principales limitaciones para ello es el principio de **localidad**. Hemos hablado un poco de la localidad en el contexto del EPR y la medición cuántica, pero ahora solamente pensamos en el comportamiento clásico de los campos. Ese comportamiento es totalmente local, no tiene nada de fantasmal.

Para un campo, la localidad significa que su evolución en un punto cualquiera del espacio-tiempo solamente depende del valor de ese campo y de otros campos en ese mismo punto, así como en la vecindad inmediata de ese punto. «Vecindad inmediata» es una forma de decir que podemos apelar a las *derivadas* del campo en el espacio o en el tiempo (ya que reflejan lo que el campo está haciendo infinitesimalmente cerca), pero el comportamiento del campo en (x,t) no depende directamente de nada que ocurra en ningún otro punto situado a una distancia finita. En la práctica, esto significa que el comportamiento de un campo ϕ en el punto (x,t) puede verse afectado por el propio valor del campo $\phi(x,t)$, así como por sus derivadas parciales con respecto a espacio y al tiempo:

$$\frac{\partial \phi}{\partial x}, \qquad \frac{\partial \phi}{\partial t}.$$

Recuerde que la derivada parcial solamente dice «toma la velocidad de cambio con respecto a esta variable en particular,

manteniendo todas las demás variables fijas». Así, en cada punto (x,t) del espacio y del tiempo, el campo tendrá derivadas espaciales y una derivada temporal. Estas derivadas aparecen tan a menudo, y los teóricos de campos se cansan tanto de escribir fórmulas imposiblemente largas, que inevitablemente utilizan una notación abreviada:

$$\frac{\partial\phi}{\partial x}=\partial_x\phi, \qquad \frac{\partial\phi}{\partial t}=\partial_t\phi\,. \tag{4.1}$$

Se supone que se sabe que estas cantidades son funciones de (x,t), aunque la dependencia no esté escrita explícitamente. Y la derivada espacial $\partial_x\phi$ junto a la derivada temporal $\partial_t\phi$, así como el propio valor del campo ϕ, son lo que podemos utilizar para construir una expresión para la energía del campo.

Otra consecuencia de la localidad es que la energía total en una configuración de campo (que, al fin y al cabo, se extiende por todo el espacio) puede considerarse como la integral en el espacio de una **densidad de energía**, la cantidad de energía por volumen de espacio, definida en cada punto. La densidad de energía se denota tradicionalmente con la letra griega ρ (rho). La energía total en una configuración de campo es entonces

$$E = \int \rho(x)\, dx. \tag{4.2}$$

De nuevo, en realidad es una integral en todo el espacio tridimensional, con el elemento de volumen $d^3x = dx\, dy\, dz$, pero por comodidad de notación escribimos como si estuviéramos en una sola dimensión.

Por último, damos la expresión para la densidad de energía $\rho(x,t)$ de un campo escalar:

$$\rho(x,t) = \frac{1}{2}(\partial_t \phi)^2 + \frac{1}{2}(\partial_x \phi)^2 + V(\phi) \qquad (4.3)$$

densidad de energía = cinética + gradiente + potencial.

Existe un parecido evidente con la energía de una partícula. Al igual que en el caso de la partícula, existe una energía cinética que depende de «cómo de rápido se mueva el sistema», salvo que ahora se trata de cómo de rápido cambia el valor del campo en un punto concreto del espacio (como una lámina de caucho que vibra), en lugar de cómo de rápido se desplace una partícula por el espacio. También existe una energía potencial que depende de la configuración instantánea del sistema, pero de nuevo se trata del valor del campo y no de la ubicación de una partícula. Lo nuevo en el caso del campo es la **energía de gradiente** ½ $(\partial_x \phi)^2$. Esto refleja el hecho de que se necesita energía para que el campo se doble a medida que nos desplazamos de un punto a otro del espacio, del mismo modo que se necesita energía para que el campo evolucione con el tiempo.

Todo esto debería tener sentido de forma intuitiva; el comportamiento de un campo clásico no es muy diferente del comportamiento de una lámina de caucho estirada. Hay energía cinética de su movimiento hacia arriba y hacia abajo en cada punto, energía de gradiente debida al estiramiento a medida que cambia de un lugar a otro, y energía potencial simplemente del valor del campo al desplazarse respecto a su valor de equilibrio en reposo.

Cabe citar aquí una característica de la fórmula de densidad de energía (4.3) que podría plantear preguntas. Tanto la energía cinética como la energía de gradiente tienen el número ½ multiplicándolas. ¿Cómo sabemos que es exactamente ese número? ¿Y por qué el número es el mismo para ambos términos?

La respuesta es la relatividad. La energía cinética aplica derivadas con respecto al tiempo, y la energía de gradiente aplica derivadas con respecto al espacio. Le puede preocupar que estas cantidades tengan

diferentes unidades, pero implícitamente estamos definiendo la velocidad de la luz como la unidad, $c = 1$, de modo que el tiempo y el espacio puedan compararse en términos conmensurables. La razón por la que los números que multiplican las derivadas del tiempo y del espacio son los mismos es que queremos que nuestra teoría obedezca las reglas de la relatividad, es decir, que sea **invariante de Lorentz**, o que tenga el mismo aspecto en todos los sistemas de referencia. Y para que eso ocurra, en unidades en las que $c = 1$, los coeficientes del espacio y el tiempo deben ser iguales. La razón por la que es ½ es algo así como una convención, pero en última instancia es el mismo factor que aparece en la expresión $\frac{1}{2} mv^2$ para la energía cinética de una partícula.

Es completamente posible inventar teorías de campo no invariantes de Lorentz; podrían ser relevantes, por ejemplo, si queremos hablar del movimiento de los electrones a través de un metal. En ese caso, no todos los sistemas de referencia se crean del mismo modo: el sistema en reposo del metal es especial. Pero en este libro nuestro interés va a estar en el comportamiento fundamental de los campos por sí mismos, no en sus propiedades concretas cuando estén inmersos en un entorno específico, de forma que obedezcan las reglas de la relatividad como deseamos.

CAMPOS LIBRES

No hay mucho margen para jugar en la expresión (4.3) para la densidad de energía. Lo único que queda sin determinar es la energía potencial $V(\phi)$. De hecho, gran parte de la emoción y la riqueza de la física de partículas va a surgir de la elección adecuada del potencial, no solamente para un único campo, sino para múltiples campos que interaccionan entre sí. En el mundo real, cada punto del espacio-tiempo presenta una ruidosa cacofonía de campos que se dan

empujones mutuamente de formas especificadas con precisión por la teoría.

Pero empezaremos suavemente, con un solo campo y el potencial no trivial más simple que se nos ocurre:

$$V(\phi)=\frac{1}{2}m^2\phi^2. \tag{4.4}$$

El parámetro m se conoce como la **masa** del campo. Los campos no tienen realmente «masa», por supuesto, ya que la masa es la energía inherente a un objeto localizado en reposo, y los campos no son objetos localizados, sino que están repartidos por todo el espacio. Pero al cuantizar descubriremos que nuestra teoría de campos describe partículas, y m será la masa de esas partículas. Por comodidad, nos referiremos a ella como la «masa del campo».

¿Por qué es este el potencial más simple? La justificación es paralela a nuestra discusión en *Espacio, tiempo y movimiento* sobre por qué el oscilador armónico simple es tan omnipresente. Piense en elevar la variable de campo a distintas potencias: ϕ^0, ϕ^1, ϕ^2, ϕ^3 y así sucesivamente. ϕ^0 es solamente una constante; añadirla a la densidad de energía no cambia la dinámica del campo, que solamente se preocupa de cómo cambia la energía de un lugar a otro o de un momento a otro. ϕ^1 es un potencial lineal que empujaría el campo en una dirección u otra. Pero queremos considerar campos a los que se les permita permanecer en un estado de energía mínima, en lugar de alejarse de él inmediatamente. Así que descartamos ese término por hipótesis.[1] ϕ^2 es algo perfectamente razonable para tener en cuenta, así que empezamos por ahí. No necesitamos preocuparnos por ϕ^3 y términos de orden superior, al menos por el momento.

Un campo cuyo potencial adopta la forma simple (4.5) se denomina **campo libre**, porque se propaga libremente, sin interaccionar consigo mismo ni con otros campos. De hecho, es el único caso en la TCC en el que realmente podemos resolver todas las

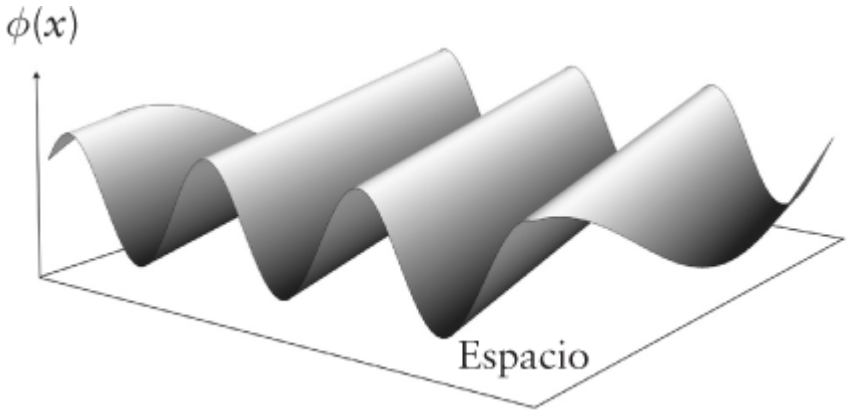
ecuaciones con exactitud. Potenciales más complicados generan interacciones entre campos, que dejaremos para el próximo capítulo.

MODOS

Ya casi estamos listos para cuantizar nuestros campos. Pero antes de hacerlo, vamos a adelantar un problema. Cuando consideramos previamente la mecánica cuántica de una partícula individual con posición x, construimos una función de onda $\Psi(x)$, signando una amplitud compleja a cada posición. Ahora vamos a tener una función de onda de configuraciones de campo, $\Psi[\{\phi(x)\}]$, asignando una amplitud compleja a cada perfil posible a través del espacio. Esto parece bastante difícil de manejar. ¿Cómo vamos a reducirlo a algo que podamos interpretar fácilmente o, para el caso, que nos permita hacer cálculos con facilidad?

La respuesta es pensar en configuraciones de campo especialmente sencillas: las **ondas planas**, vibraciones que parecen ondas oscilatorias perfectamente regulares que viajan por todo el espacio. Una onda plana es tan simple que podemos describir su comportamiento con exactitud.

Pero las ondas planas no tendrían mucho interés si solamente fueran un tipo muy especial de configuración de campo fácil de analizar. Su verdadero poder reside en un hecho asombroso: cualquier configuración de campo puede considerarse una combinación de diferentes tipos de ondas planas.



Esa es la magia de las transformadas de Fourier, el truco matemático que mencionamos en el capítulo 2 y que analizamos en detalle en el apéndice. Vamos a ponernos un poco específicos matemáticamente, pero merecerá la pena el esfuerzo.

Así, en lugar de pensar que una configuración de campo está definida por un valor separado en cada punto del espacio $\{\phi(x)\}$, podemos pensar que es una suma de ondas oscilantes que se extienden por todo el espacio, cada una con una longitud de onda y una dirección de movimiento específicas. Las ondas individuales que sumamos se denominan **modos**. El primer truco (de muchos) que se utiliza para dar sentido a los campos cuánticos es pensar modo por modo, en lugar de pensar punto por punto en el espacio. En la práctica, en lugar de tratar con la función de onda de todo el campo, trataremos con la función de onda de un único modo, y luego juntaremos muchos de esos modos. Y cuando nos enfoquemos de cerca en el comportamiento de un solo modo, va a ser simplemente un problema de mecánica cuántica normal. (De hecho, va a ser un oscilador armónico simple).

Estamos considerando un campo escalar libre, con energía dada por (4.3) y potencial dado por (4.4), y vamos a pensar en la energía de un único modo con longitud de onda fija antes de combinar muchos modos para crear la configuración que queramos. Si la longitud de onda es λ, lo más conveniente es caracterizar nuestro modo en

términos de su **número de onda** k, que es 2π veces el número de longitudes de onda por distancia:

$$k = 2\pi/\lambda. \tag{4.5}$$

Esto es conveniente porque podemos escribir la configuración de campo de un modo particular como la suma de una parte real oscilante y una parte imaginaria oscilante, como

$$\tilde{\phi}(k,t) = a(t)\cos(kx) - ia(t)\sin(kx) = a(t)e^{-ikx}, \tag{4.6}$$

donde $a(t)$ representa la amplitud de esta onda vibratoria en particular. Véase el apéndice sobre las transformadas de Fourier para más discusión, incluyendo la relación (A.4) entre las funciones trigonométricas y la exponencial utilizando un número imaginario. Esquemáticamente (escribiendo una suma en lugar de una integral, para mantener las cosas conceptualmente claras), podemos combinar modos en cada número de onda k, con amplitudes específicas a_k para cada modo, a fin de describir cualquier configuración de campo:

$$\phi(x,t) = \sum_k a_k(t)e^{-ikx}. \tag{4.7}$$

Hay algunos detalles en los que no vamos a detenernos ahora. Sí destacaremos una función de amplitud separada para cada modo, razón por la que hemos escrito la función de amplitud en (4.7) como $a_k(t)$. Pero hay tantos subíndices por aquí y por allá que suprimiremos este de ahora en adelante y nos limitaremos a escribir $a(t)$. Clásicamente, podemos resolver para la forma real de $a(t)$, que oscila en el tiempo con una frecuencia que está fijada por las propiedades del campo. Por último, en tres dimensiones, hablaríamos del **vector de onda** $\vec{k}$, cuyo número de onda es su magnitud: $k = \left|\vec{k}\right|$. Entonces un

modo tendría la forma $\phi\left(\vec{k}\right)=a_{\vec{k}}(t)e^{-i\vec{k}\cdot\vec{x}}$, donde $\vec{k}\cdot\vec{x}$ es el producto escalar del vector de onda y la posición. Pero continuaremos simplificando nuestras vidas fingiendo que el espacio solo tiene una dimensión, al menos en lo que se refiere a nuestra notación, por lo que cada uno de nuestros modos tendrá una dependencia espacial e^{-ikx}.

La razón por la que la teoría cuántica de campos puede parecer intimidatorio, incluso si se han hecho las paces con los fundamentos de la mecánica cuántica, es que hay un número infinito de cosas que tener en cuenta en la función de onda de las configuraciones de campo, $\Psi[\{\phi(x)\}]$. Existen muchas configuraciones de campo ahí fuera. ¿Cómo podemos seguir la pista a todas ellas? Podríamos pensar que podemos centrarnos en un único punto del espacio x_0, estudiar el comportamiento del campo cuántico $\phi(x_0)$ en ese punto y luego generalizarlo a todos los puntos del espacio. Pero eso no funciona del todo: debido a los términos de la derivada espacial (la energía de gradiente) en (4.3), lo que ocurre en un punto depende de lo que ocurre en todos los puntos cercanos de manera no trivial.

Los modos vienen en nuestra ayuda. El poder de la transformada de Fourier es que podemos observar un modo cada vez, en lugar de una ubicación espacial cada vez. Y en ese caso la dependencia espacial del campo está completamente resuelta: es e^{-ikx}, como se indica en (4.6). Así que solamente tenemos que concentrarnos en la función de onda de un único número, la amplitud a. Eso no debería ser más difícil que en la mecánica cuántica ordinaria, cuando nos interesábamos por la función de onda de una única posición x. En ambos casos tenemos un único número (posición de la partícula, amplitud del modo) del que dependerá nuestra función de onda. Para partículas tenemos $\Psi(x)$, para un único modo de un campo libre tendremos $\Psi(a)$, pero las manipulaciones serán muy parecidas.

LA ENERGÍA DE UN MODO

Introduzcamos ahora la forma explícita (4.6) para un único modo de nuestro campo en las ecuaciones (4.3) y (4.4) para su densidad de energía. Observe en primer lugar que la derivada de una función exponencial hace descender lo que sea que esté multiplicando a la variable en el exponente, por lo que

$$\frac{\partial}{\partial x} e^{-ikx} = -ik\ e^{-ikx}. \tag{4.8}$$

El otro factor que hay que tener en cuenta es que, después de la transformada de Fourier, todas las cantidades al cuadrado de la densidad de energía (4.3) se convierten en valores absolutos al cuadrado. Así que todos esos exponenciales sinusoidales desaparecen, por ejemplo:

$$\left|e^{ikx}\right|^2 = e^{ikx} \cdot e^{-ikx} = 1. \tag{4.9}$$

En definitiva, la densidad de energía del modo (4.6) tiene el aspecto siguiente:

$$\begin{aligned} \tilde{\rho}(k) &= \frac{1}{2}\left(\partial_t a\right)^2 + \frac{1}{2}a^2k^2 + \frac{1}{2}a^2m^2 \\ &= \frac{1}{2}\left(\frac{da}{dt}\right)^2 + \frac{1}{2}\left(k^2 + m^2\right)a^2. \end{aligned} \tag{4.10}$$

(Podemos sustituir las derivadas parciales del primer término por una derivada ordinaria, ya que a solamente depende de una variable, t). Cuando decimos «densidad» de energía, ahora estamos hablando de la energía por modo k, no por localización en el espacio x. Los físicos suelen hablar de describir campos «en el espacio de Fourier», en lugar de «en el espacio real». Esto significa que estamos pensando

en el campo como una colección de modos de ondas planas en lugar de como una colección de valores en puntos del espacio ordinario. Son completamente equivalentes, pero la descripción como modos es más conveniente para cuantizar la teoría.

Bueno, no deje que ahora se le nublen los ojos. El pequeño cálculo que acabamos de hacer para derivar (4.10) para la densidad de energía de un modo con número de onda k es una de las ideas más importantes de la teoría cuántica de campos y, por tanto, una de las más notorias de la ciencia moderna. Quizá incluso de la historia de la humanidad.

Para ver por qué, olvidemos por un momento todo lo que acabamos de decir sobre la TCC y pensemos en una sola partícula en un potencial de oscilador armónico simple. Su energía es

$$E_{\mathrm{OAS}} = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\omega^2 x^2. \tag{4.11}$$

Observe esto un momento. Ahora observe la ecuación (4.9). Luego vuelva a observar aquí.

Lo que ve es que son la misma ecuación. La única diferencia es algún reetiquetado. En lugar de la posición de la partícula $x(t)$, tenemos la amplitud del modo oscilante, $a(t)$. Y el papel de la frecuencia angular w lo desempeña ahora $\sqrt{k^2 + m^2}$. Pero la estructura matemática es la misma. Esta es la idea crucial: si tomamos un campo escalar libre y lo descomponemos en un conjunto de modos de número de onda k, cada modo individual actúa exactamente como un oscilador armónico simple de frecuencia

$$\omega^2 = k^2 + m^2. \tag{4.12}$$

La frecuencia con la que oscilará un modo depende tanto de su número de onda como de la masa del campo. Un campo cuya masa es igual a cero puede oscilar arbitrariamente despacio (cuando k es muy

pequeño), pero la masa impone un límite inferior a la frecuencia. Y, a medida que el número de onda aumenta (lo que corresponde a longitudes de onda más pequeñas), la frecuencia aumenta.

Este tipo de prestidigitación matemática es habitual en física. Conocemos un problema sencillo: una sola partícula en un potencial de oscilador armónico. En el fondo de nuestra mente podría haber un sistema físico tangible, tal vez un electrón atrapado en un campo eléctrico, o tal vez un peso colgando de un muelle. Pero a las ecuaciones no les importa en qué sistema estemos pensando. Representan la esencia de la dinámica reducida a su forma más básica, que puede ser común a muchos sistemas distintos. A continuación, pasamos a una situación física completamente distinta: una configuración de onda plana de un campo escalar libre, que oscila hacia arriba y hacia abajo en el tiempo. Sorprendentemente, las ecuaciones resultan ser las mismas (con algún sencillo reetiquetado). Así que nuestro trabajo ya está hecho; ya sabemos qué aspecto tienen las soluciones. En este caso, volvemos a encontrar niveles de energía cuantizados.

Si (4.12) le resulta familiar, recuerde la discusión sobre energía y momento relativistas en *Espacio, tiempo y movimiento*. Allí descubrimos que la energía, el momento y la masa de una partícula se relacionan mediante la ecuación $E^2 = p^2 + m^2$. Esto no es una coincidencia. Una vez que sacamos algunas partículas de nuestros campos cuánticos, la energía de la partícula asociada a este tipo de modo será $E = \hbar w$ y el momento será $p = \hbar k$, donde $\hbar$ es la constante de Planck (que ha sido invisible desde que la fijamos, y la velocidad de la luz, igual a 1). Se trata de una serie de relaciones generales que conviene tener en cuenta. Recordando que los físicos etiquetan como ultravioleta o «*UV*» a los fenómenos a corta distancia e infrarrojo o «*IR*» a los fenómenos a larga distancia, tenemos las siguientes correspondencias:

UV	IR

Longitud de onda corta	Longitud de onda larga
Tiempos cortos	Tiempos largos
Alta energía	Baja energía

La moraleja de esta historia es que un campo escalar libre puede considerarse una colección infinita de osciladores armónicos. No es que haya un oscilador en cada punto del espacio; los «osciladores» son modos individuales que se extienden por el espacio y pueden sumarse para describir cualquier configuración de campo que queramos. Ya le conté que el oscilador armónico simple era la mejor vaca esférica que existe.

FUNCIONES DE ONDA DE LOS CAMPOS

Ahora estamos listos para cuantizar nuestros campos. Por razones de claridad conceptual, plantearemos el problema en términos de configuraciones de campo $\{\phi(x)\}$, y en la siguiente sección pasaremos a utilizar modos para hacernos la vida más fácil.

Cuando hablamos del entrelazamiento, hicimos hincapié en que la función de onda no es un campo, al menos cuando hay más de una partícula en el universo. Los campos son cosas que tienen un valor definido asociado a cada punto del espacio, mientras que las amplitudes de la función de onda están asociadas a configuraciones de cualquier sistema físico que nos podamos imaginar. Si ese sistema consta de varias partículas, el espacio de configuración consiste en varias copias del espacio físico, una para cada partícula, así que tenemos $\Psi(x_1, x_2, \ldots x_N)$ para N partículas.

Si la función de onda no es un campo, pero estamos haciendo teoría cuántica de campos, ¿cuál es la relación entre ambas? La respuesta es que tenemos una función de onda *de* configuraciones de campo, exactamente en el mismo sentido que una teoría cuántica de partículas presenta una función de onda de posiciones de partículas.

Es decir, en una teoría de partículas tomamos el conjunto de todas las posibles ubicaciones de una partícula (o de un conjunto de varias partículas), y a cada una le asignamos un número complejo. El conjunto de todas estas asignaciones es la función de onda, que define el estado cuántico. En una teoría de campos, tomamos el conjunto de todas las configuraciones posibles del campo, y a cada configuración completa le asignamos un número complejo. La función de onda del campo es el conjunto de todas esas asignaciones, que podemos escribir como $\Psi[\{\phi(x)\}]$. Nótese que $\Psi[\{\phi(x)\}]$ no es una función de x, sino de toda la configuración $\{\phi(x)\}$, que está definida en todos los puntos del espacio. La notación es complicada, pero lo haremos lo mejor que podamos.

Este sinsentido abstracto es un poco más fácil de digerir en el caso de las partículas porque podemos imaginarnos observando la partícula en distintas posiciones del espacio, y la probabilidad de obtener ese resultado de la medición es la función de onda correspondiente elevada al cuadrado. En el caso del campo, tenemos que tirar un poco de nuestra imaginación y pensar que estamos midiendo simultáneamente el campo en cada punto del espacio. No es práctico desde el punto de vista realista, pero nos ayuda a definir de qué estamos hablando. La probabilidad de observar una determinada configuración de campo es

$$P[\phi(x)] = \left|\Psi\left[\{\phi(x)\}\right]\right|^2. \tag{4.13}$$

De nuevo, esta no es la probabilidad de medir $\phi(x)$ de tener un valor concreto en algún punto concreto x; es la probabilidad de tener una configuración entera $\{\phi(x)\}$ para todo el espacio.

También existe una versión de la ecuación de Schrödinger que nos dice cómo evoluciona la función de onda con el tiempo. Todo lo que tenemos que hacer es escribir el operador hamiltoniano apropiado que caracteriza la energía del estado, y entonces podemos resolver $\widehat{H}\Psi = i\hbar\partial\Psi/\partial t$. Pero así como existen alternativas equivalentes a la formulación newtoniana de la mecánica clásica (mecánica hamiltoniana, mecánica lagrangiana), también existen formas alternativas de caracterizar las funciones de onda de los estados cuánticos, que resultarán más útiles para la TCC.

PARTÍCULAS DE LOS CAMPOS

Ahora que ya hemos hablado en abstracto de lo que se supone que significa una «función de onda de un campo», podemos ir al grano. La buena noticia es que una vez que cambiamos las variables de configuraciones de campo $\{\phi(x)\}$ a modos de onda plana $\{\phi(k)\}$, cuantizar nuestro campo es realmente un juego de niños. La respuesta va a ser que los cuantos que surgen de campos pueden interpretarse como conjuntos de partículas.

El campo libre, que podemos considerar una suma de modos, es una colección de osciladores armónicos simples, uno para cada modo con un número de onda k dado. Piense en los modos de uno en uno. Esencialmente nos interesa la función de onda de la variable de amplitud única a para cada modo k, que acabamos de establecer que se comporta como un oscilador armónico. La función de onda de este modo, $\Psi(a)$, tendrá por tanto el conjunto característico de soluciones que vimos para el oscilador en el Capítulo 1. Hay un **estado fundamental** de energía mínima y luego una pila de **estados excitados** con energías uniformemente espaciadas por encima de la energía del estado fundamental.

El conjunto de estados cuánticos de toda la TCC no es más que la colección de todos estos estados de osciladores armónicos para cada

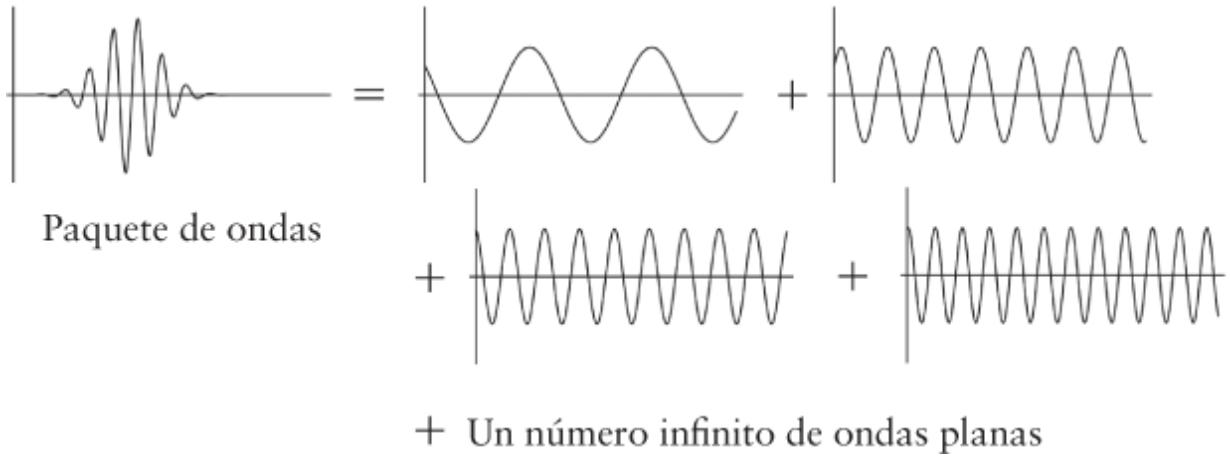
vector de onda k. Habrá un único estado de mínima energía global, aquel en el que todos los osciladores individuales se encuentren en sus estados fundamentales. Es el llamado **estado de vacío** de la teoría. Los físicos no utilizan «vacío» para referirse a «espacio sin contenido» (aunque eso es lo que parecerá), sino al «estado de menor energía de una teoría».

Únicamente hay una forma de tener energía cero, pero hay muchas formas de tener energía distinta de cero. Empezando por el vacío, podemos imaginar «dar vida» a un solo modo con número de onda k. Dado que los modos son como osciladores armónicos, el estado cuántico puede considerarse una superposición de los niveles de energía de ese modo. Y un estado cuántico más general del campo en su conjunto será una superposición de dichos estados para todo el número infinito de modos de varios k. Hay que hacer muchos malabarismos, pero podemos ir desarrollando la idea poco a poco hasta pillar el truco a lo que está pasando.

Piense en un estado en el que solamente intervenga un modo, el modo con $k = 0$. Dado que $k = 2\pi/\lambda$, se trata de un modo con «longitud de onda infinita», que quiere decir que, básicamente, el campo tiene un valor constante en todas partes. Desde el punto de vista clásico, este valor puede oscilar con el tiempo; desde el punto de vista de la mecánica cuántica, tendrá una función de onda con cierto perfil $\Psi(a)$. Imaginemos que este modo se encuentra en su primer estado excitado. A partir de (4.12) la frecuencia del potencial del oscilador armónico asociado es simplemente $w = m$. La energía del estado es, por tanto, $E = m$, ya que $E = w$, y el momento es $p = k = 0$. (Recuerde que $\hbar = 1$). Eso es interesante: son justamente la energía y el momento de una sola partícula de masa m en reposo.[2]

Consideremos ahora un único modo con algún otro número de onda k, también en su primer estado excitado. Tendrá una energía $E = \omega = \sqrt{k^2 + m^2}$, exactamente como esperamos para una única partícula de momento $p = k$. Su perfil espacial no es constante, sino que tiene el aspecto de una onda plana, e^{-ikx}.

¿Y si estamos interesados en configuraciones de campo que no son tan simples como un único modo de onda plana? No hay problema: podemos combinar los modos de diferentes k, todos ellos en sus primeros estados excitados, para obtener cualquier perfil de campo que queramos. Una transformada de Fourier toma un perfil de campo arbitrario y lo expresa como una suma de ondas planas, pero también podemos ir hacia atrás, combinando una colección de ondas planas para que tengan cualquier forma que nos interese. Así, por ejemplo, podemos crear un **paquete de ondas**, que es como los físicos llaman a una onda localizada que oscila cerca de una posición determinada en el espacio, pero que tiende a 0 lejos de ella.



Lo que vemos es que un estado cuántico construido a partir de una superposición de modos de campo libre en sus primeros estados excitados se parece a la función de onda de una sola partícula de masa m y actúa como ella. No hemos agregado partículas: hemos empezado con un campo y lo hemos cuantizado. Las soluciones de las ecuaciones resultantes son cuantos parecidos a partículas. Este es un caso en el que si algo parece un pato y grazna como un pato, podemos declarar justificadamente que es un pato. El primer estado excitado de un campo cuántico libre puede interpretarse como el estado cuántico de una única partícula relativista.

Al describir la energía del campo en (4.3), introdujimos el potencial $V(\phi)$, pero esa es una forma de pensar en la energía que está asociada al valor del campo (en lugar de a sus derivadas). La partícula que ha surgido tras la cuantización del campo no está atrapada en ningún tipo de potencial; se mueve libremente por el espacio. En el límite clásico, se movería a velocidad constante en línea recta. Es la amplitud del modo cuya energía implica un potencial de oscilador armónico y, por tanto, tiene niveles de energía discretos. Estamos haciendo un uso inteligente de una equivalencia matemática entre la energía potencial de un oscilador armónico real y la correspondiente a la amplitud de un modo de campo.

El segundo estado excitado de nuestro campo cuántico libre —hecho a partir de la combinación de los segundos estados excitados de varios modos— tendrá una energía mínima de $2m$. El tercer estado excitado tendrá una energía mínima de $3m$, y así sucesivamente. Esto se debe a que corresponden a estados que parecen dos partículas, tres partículas, y así sucesivamente.

Observar nuestra TCC en términos de modos ha revelado algo asombroso: un estado de una TCC libre puede interpretarse como una superposición de colecciones de diferentes números de partículas. Podemos pensar en ello en términos del espacio de Hilbert de la TCC:

$$\text{TCC Espacio de Hilbert} = \left(\begin{array}{c}\text{Estado único de cero partículas}\end{array}\right) + \left(\begin{array}{c}\text{Estado de una partícula}\end{array}\right) + \left(\begin{array}{c}\text{Estado de dos partículas}\end{array}\right) + \cdots$$

Esta descripción se conoce en la literatura como **espacio de Fock**, en honor al físico ruso Vladimir Fock. No se trata de algo nuevo, sino solo de una forma agradable de concebir el espacio de Hilbert de una teoría cuántica de campos.

Se podría pensar que ser una «partícula» es algo más que tener niveles de energía, y en cierto sentido es correcto. Más concretamente, estamos acostumbrados a que las partículas tengan localizaciones

definidas en el espacio. Pero esta intuición procede de la visión clásica de las partículas. En mecánica cuántica, las partículas tienen funciones de onda, y la localización en el espacio solamente se produce cuando medimos su posición. Lo mismo ocurre con las partículas descritas por nuestro campo cuántico. Tendrán funciones de onda. Es muy probable que esas funciones de onda estén repartidas por el espacio, pero también somos libres de considerar paquetes de ondas relativamente compactos centrados en alguna localización. La cuestión es que los tipos de partículas que surgen de nuestros campos cuantizados tienen todas las propiedades y comportamientos que se supone que tienen las partículas cuánticas.

Todo esto es bastante sorprendente. En el oscilador armónico original (o las órbitas de los electrones en un átomo, para el caso), la naturaleza «cuántica» de la mecánica cuántica no se concedió a dedo. Surgió de forma natural al resolver la ecuación de Schrödinger y encontrar un conjunto discreto de soluciones. Aquí ocurre exactamente lo mismo. Hemos cuantizado un campo, no un conjunto de partículas, pero hemos descubierto que acaba pareciéndose a una colección de cuantos, que interpretamos como partículas. Y no es solamente semántica: esto es lo que son realmente las partículas: excitaciones cuantizadas de campos. Esta es la recompensa que obtenemos por pensar detenidamente en los modos y sus energías.

CREACIÓN Y ANIQUILACIÓN

Para ser justos, la vaca que estudiamos es realmente muy esférica. Una teoría de campo libre —aquella cuya energía es la suma de los términos de energía cinética, de gradiente y potencial como en la ecuación (4.3), con el potencial cuadrático simple (4.4)— es un buen punto de partida porque podemos resolver todas las ecuaciones con exactitud, pero da lugar a un universo muy poco interesante. Las partículas de una teoría de campo libre no pueden irradiar,

desintegrarse, dispersarse, ni tampoco convertirse una en otra ni interaccionar entre ellas. Para describir cualquiera de esos importantes procesos, tendremos que complicar nuestra vaca metafórica añadiendo interacciones, por no mencionar la introducción de más de un tipo de campo.

Pero podemos ver, incluso a partir de nuestra investigación hasta ahora, que la teoría de campos nos da los ingredientes para describir partículas que se crean y se destruyen, como vemos en la radiactividad. No era obvio cómo hacer esto en una teoría cuántica que se basaba en partículas desde el principio, en lugar de en campos. Si tengo un estado cuántico de, digamos, dos partículas, la función de onda puede evolucionar de muchas maneras, pero siempre describirá dos partículas, mientras que los estados cuánticos de un solo campo incluyen estados con cero partículas, una partícula, dos partículas, etcétera. Así pues, la creación y aniquilación de partículas queda totalmente inmersa en la dinámica de un solo campo cuántico (o de varios).

Si tuviéramos alguna aversión profundamente arraigada a la teoría cuántica de campos, podríamos intentar permitir procesos que cambiaran el número de partículas dentro de un enfoque basado en partículas. Para ello tendríamos que aumentar el espacio de Hilbert, sumando los espacios apropiados para cero partículas, una partícula, dos partículas, etcétera. Luego tendríamos que modificar el hamiltoniano de alguna manera para permitir transiciones entre estas piezas, dejando (por ejemplo) que una sola partícula se desintegrase en varias partículas. (En el mundo real, un neutrón abandonado a su suerte se desintegrará en un protón más un electrón más un antineutrino). En otras palabras, tendríamos que inventar el espacio de Fock de todos modos. Haciendo una investigación más profunda, descubriríamos que en realidad habíamos recreado toda la teoría cuántica de campos, aunque con una serie de movimientos especialmente cargada de rodeos.

Nos enfrentamos a lo que los físicos llaman a veces un **teorema popular**: un resultado que se considera cierto en una gran variedad de circunstancias, pero que no puede demostrarse de forma rigurosa. El teorema popular en cuestión es el siguiente: si quieres satisfacer simultáneamente los requisitos de (1) mecánica cuántica, (2) relatividad especial, (3) localidad, y (4) cambiar el número de partículas, entonces cualquier teoría que construyas acabará pareciéndose a la teoría cuántica de campos, al menos, a bajas energías/largas distancias. (A distancias cortas podríamos descubrir la teoría de cuerdas o dimensiones extra o alguna otra cosa exótica). La TCC, en otras palabras, es increíblemente sólida: es una forma prácticamente exclusiva de satisfacer un conjunto de requisitos sencillos.

CAMPOS DEL MUNDO

La teoría cuántica de campos es un tema rico y complejo. Muchas personas inteligentes dedican su vida a entenderla (y, sinceramente, hay formas menos gratificantes de pasar la vida). Existen muchos aspectos que contribuyen a esta riqueza. Destaquemos solamente dos.

En primer lugar, hay un montón de campos. Cada uno de ellos tiene su propia idiosincrasia especial. Consideraremos dos grandes clases: los **fermiones**, que corresponden aproximadamente a las partículas de «materia», y los **bosones**, que crean los campos de fuerza que conocemos. Los fermiones incluyen los electrones, los quarks y los neutrinos, entre otros; los bosones incluyen los fotones, los gluones, los gravitones, el bosón de Higgs, etcétera.[3] Los diferentes campos tienen espines distintos, se caracterizan por simetrías diferentes y, por tanto, tienen sus propias notaciones. Hay de todo en la colección.

En segundo lugar, los campos interaccionan de mil maneras, y calcular el efecto físico de esas interacciones puede ser un poco lioso. Podríamos decir que «dos electrones se dispersan entre sí», pero lo que

realmente ocurre es que las vibraciones de algunos campos apropiados empiezan a solaparse, lo que induce vibraciones en otros campos a los que están acoplados, y literalmente tenemos una cantidad infinita de actividad contribuyendo al más simple de los procesos. Ahora que hemos establecido que la cuantización de los campos conduce hasta las partículas, en adelante dedicaremos nuestro tiempo a explicar cómo pensar en sus interacciones. Al fin y al cabo, estas interacciones son las que hacen del universo un lugar interesante.

---

1 Con algo más de detenimiento: imaginemos que el potencial tuviera un mínimo en algún valor $\phi_{\text{min}} \neq 0$. Podríamos simplemente definir una nueva variable, $\phi' = \phi - \phi_{\text{min}}$. Entonces el mínimo estaría en $\phi' = 0$, así que sabemos que no habría ningún término $(\phi')^1$ en el potencial.

2 Detalle molesto: la fórmula (1.14) para los niveles de energía de un oscilador armónico simple incluye una contribución constante $1/2\ \hbar\omega$. Se trata de la energía del punto cero, y vamos a ignorarla por ahora. No pasa nada, porque lo que importa físicamente son las diferencias de energía entre los distintos estados, no la energía global. Pero cuando pensamos en la atracción gravitatoria, esta energía de punto cero del campo cuántico se convierte en la «constante cosmológica»; es un tema que tocaremos brevemente más adelante.

3 El gravitón no se ha detectado directamente, y es posible que nunca se detecte, porque los gravitones individuales interaccionan muy débilmente con otras partículas. Únicamente notamos la gravedad de la Tierra porque sentimos la atracción gravitatoria combinada de 10^{50} átomos. Pero los principios básicos de la teoría cuántica de campos y la relatividad general prácticamente garantizan la existencia de los gravitones. Su particularidad es que es una partícula de espín 2.

CINCO

INTERACCIONES

Fue toda una revelación la que tuvimos en el último capítulo. Empezamos con la teoría de campos clásica más simple posible —un único campo escalar $\phi(x, t)$, que no interacciona con ningún otro campo ni consigo mismo, con un potencial simple $V(\phi) = \frac{1}{2} m^2\phi^2$— y lo introducimos en las reglas de la mecánica cuántica. Lo que encontramos es un conjunto de estados con energías diferentes, que parecen y se comportan exactamente como conjuntos de partículas. Hay un estado de vacío, estados con una sola partícula, estados con dos partículas, etcétera. La teoría clásica original era todo lo continua, uniforme y no discreta que podíamos imaginar, pero al trasladarla a un marco cuántico se convierte en una teoría de cuantos que identificamos como partículas. Y, al igual que el campo del que proceden, esas partículas no van a interaccionar: si establecemos una configuración inicial que describa varias partículas moviéndose unas hacia otras, cada una de ellas va a continuar moviéndose en línea recta sin verse afectada por ninguna otra partícula.

El mecanismo subyacente para la aparición de grumos a partir de la suavidad es precisamente el mismo por el que vemos niveles de energía discretos para las funciones de onda de los electrones en los átomos: cuando consideramos la ecuación de Schrödinger en un contexto apropiado, obtenemos un conjunto de soluciones discretas. El carácter cuántico de la mecánica cuántica, incluida la teoría cuántica de campos, proviene de la resolución de las ecuaciones, no de la naturaleza fundamental de los ingredientes que utilizamos para construir el modelo.

Pero no vamos a estar satisfechos con una teoría que solamente contenga un único campo no interactivo. El mundo real tiene diferentes tipos de campos y, en consecuencia, diferentes partículas, y tienden a interaccionar entre sí. Aquí es donde se manifiesta el verdadero poder de la teoría cuántica de campos: no solamente describe estados con múltiples partículas, sino también procesos en los que un conjunto de partículas evoluciona hacia otro conjunto de partículas completamente distinto. Entender qué son los campos y cómo interaccionan entre sí no solo explica en definitiva la física de partículas, sino también la tabla periódica de los elementos, las reacciones químicas y todos los peldaños que llevan al mundo macroscópico de nuestra experiencia.

ORÍGENES

A veces oiremos decir que existen cuatro fuerzas de la naturaleza: el electromagnetismo, la gravitación y las fuerzas nucleares fuerte y débil. Es una buena forma de pensar, y a veces utilizaremos esa terminología. Pero no es muy rigurosa. En el mundo de la teoría cuántica de campos, no hay una división clara entre «partículas» y «fuerzas». Solo hay campos, que interaccionan entre sí de diversas maneras.

Históricamente, la necesidad de la teoría cuántica de campos surgió porque todo el mundo conocía el campo electromagnético y reconocía que tendría que hacerse totalmente cuántico en algún momento. (También conocían la gravedad, pero la relatividad general de Einstein tiene ecuaciones más complicadas que el electromagnetismo de Maxwell, así que podríamos empezar igualmente por este último). Ya en 1925, Heisenberg, Born y Jordan estudiaron una teoría cuántica en exclusiva del campo electromagnético —sin partículas cargadas— que es esencialmente una teoría de campo libre y no es tan interesante. En 1927, Paul Dirac

descubrió cómo añadir electrones a la teoría. Esto le llevó a una investigación más cuidadosa sobre cómo describir partículas giratorias como los electrones en un contexto relativista, lo que al final condujo a su ecuación y a la predicción de la **antimateria** (positrones, las antipartículas de los electrones). Las antipartículas se denotan colocando una barra sobre el correspondiente símbolo. Así, el protón, el neutrón, el electrón y el neutrino son p, n, e y ν, mientras que sus antipartículas son $\overline{p}$, $\overline{n}$, $\overline{e}$ y $\overline{\nu}$. Toda partícula portadora de una cantidad conservada que pueda ser positiva o negativa, como la carga eléctrica, tendrá un tipo correspondiente de antipartícula con el valor opuesto de dicha cantidad. Como los electrones tienen carga –1, los positrones deben tener carga +1. Los positrones fueron descubiertos experimentalmente por Carl Anderson en 1935, pero al principio la gente no le creyó. Pensaron que había observado protones, o que simplemente había observado electrones, pero que sostenía las fotografías al revés y, por tanto, interpretaba mal cómo se curvaban las partículas en un campo magnético.

Interpretar las ideas de Dirac en términos de teoría cuántica de campos llevó algo más de tiempo. Pero algunas de las ideas básicas estaban ahí, incluida la necesidad de describir procesos en los que el número de partículas cambiaba, como cuando un electrón de un átomo desciende a un nivel de energía inferior y emite un fotón. En 1934, Enrico Fermi propuso una teoría de campos que describía otro proceso de cambio de partículas: la **desintegración beta**, en la que un neutrón se desintegra en un protón, un electrón y lo que ahora llamaríamos un antineutrino. A estas alturas, los campos cuantizados parecían un camino prometedor.

Fermi envió por primera vez su artículo sobre la desintegración beta a *Nature*, una importante revista científica inglesa, que lo rechazó por considerarlo demasiado especulativo. En parte debido a esta experiencia, Fermi cambió su enfoque de la física teórica a la experimental. Dirigió el grupo que creó la primera reacción nuclear en cadena autosostenida, que sirvió de precursora de la bomba atómica.

DISPERSIÓN

La buena noticia es que, una vez que hemos pasado por todo ese trabajo para cuantizar los campos y descubrir que se puede pensar en ellos como colecciones de partículas, podemos soltarnos el pelo y hablar simplemente como si el mundo estuviera hecho de partículas. Sabemos que en realidad está hecho de campos, y a veces será importante recordarlo, pero el *lenguaje* de las partículas es una forma útil y eficaz de hablar de la dinámica de los campos cuánticos, siempre que seamos conscientes de su ámbito de aplicación.

Hay una serie de procesos físicamente interesantes que implican interacciones entre partículas. La **dispersión** es la idea más básica: un conjunto de partículas se aproximan e interaccionan entre sí, y otro conjunto de partículas se aleja. Pueden ser las mismas partículas, como cuando dos electrones se aproximan, interaccionan mediante la fuerza electromagnética y se repelen mutuamente. Pero también podría ser diferente. Podemos considerar situaciones en las que un electrón y un positrón (su antipartícula) se aproximen y se conviertan en dos fotones, que luego irán cada uno por su lado.

Un caso especial de dispersión es la **desintegración**, que se produce cuando una partícula se convierte de manera espontánea en otra u otras. La desintegración beta de los neutrones es un ejemplo, pero hay muchísimos más. Elementos como el radio o el plutonio son radiactivos porque las combinaciones concretas de partículas que constituyen sus núcleos tienden a la desintegración. También podríamos considerar la **absorción**, en la que varias partículas se combinan para formar una sola. Este proceso es relevante en la fusión nuclear y en otros casos, pero, en general, es menos frecuente que la desintegración.

No todas las formas de interacción entre partículas son un tipo de dispersión. Cuando un electrón está unido por fuerzas electromagnéticas a un átomo, está interaccionando con el núcleo, pero no es dispersión, porque esa interacción persiste en el tiempo. Lo

mismo ocurre cuando protones y neutrones se mantienen unidos en un núcleo. La dispersión se limita específicamente a los casos en los que las interacciones pueden ignorarse al principio porque las partículas están muy separadas, luego se vuelven importantes a medida que las partículas se acercan unas a otras, y después, dado el caso, podemos volver a ignorarlas.

Esto nos facilita considerablemente la vida. La situación paradigmática que consideramos en la teoría cuántica de campos es partir de una condición inicial de partículas muy separadas, dejar que se junten e interaccionen y utilizar las leyes de la física para predecir en qué se dispersarán. Las «partículas» son en realidad vibraciones en campos cuánticos, pero no importa.

Como estamos hablando de mecánica cuántica, la predicción no será determinista. Calcularemos la amplitud para diferentes resultados de dispersión imaginables y luego elevaremos al cuadrado esa amplitud según la regla de Born para obtener la probabilidad correspondiente. (En realidad, *nosotros* no calcularemos nada, pero aprenderemos lo que se introduce en dicho cálculo). Por suerte, estos experimentos bastan para revelar una enorme cantidad de información sobre las leyes subyacentes de la física. Sin embargo, los cálculos pueden ser extremadamente difíciles de hacer, y todo un desafío de entender, incluso si no los hacemos explícitamente. Pero no estamos desanimados.

LA RECETA DE FEYNMAN

A lo largo de las décadas de 1940 y 1950 se dedicaron grandes esfuerzos al desarrollo de la teoría cuántica de campos, aunque los avances se vieron retrasados por la Segunda Guerra Mundial. Una vez terminada la guerra (y los físicos teóricos liberados en gran medida de las actividades bélicas), estos esfuerzos se redoblaron. Surgieron numerosas dificultades técnicas, centradas en el hecho de que un

cálculo ingenuo de las amplitudes básicas de dispersión parecía dar respuestas infinitamente grandes, una clara señal de que faltaba algún conocimiento básico. La historia de cómo se resolvieron estos problemas gracias a físicos como Julian Schwinger, Richard Feynman, Shin'ichirō Tomonaga y Freeman Dyson es intrincada y fascinante. Como es nuestra costumbre, vamos a pasar por alto todas las complicaciones y pasos en falso para llegar directamente a la forma más sencilla e inmediata de entender la respuesta final, en este caso, la receta desarrollada por Feynman para calcular procesos de partículas en términos de pequeños y entretenidos diagramas.

A pesar de algunos éxitos iniciales, en la década de 1940 los físicos no estaban nada convencidos de que la teoría cuántica de campos fuera por buen camino. (En algunos círculos, el escepticismo perduró hasta finales de la década de 1960). Feynman, en particular, se preguntó si no sería mejor pensar que la teoría describe en realidad muchas partículas individuales en lugar de un campo cuántico subyacente. Desarrolló una forma extremadamente útil de pensar en las interacciones en términos de partículas, aunque pronto quedó claro que la justificación subyacente de sus reglas se entendía mejor en términos de teoría cuántica de campos.

He aquí, en resumidas cuentas, el procedimiento de Feynman:

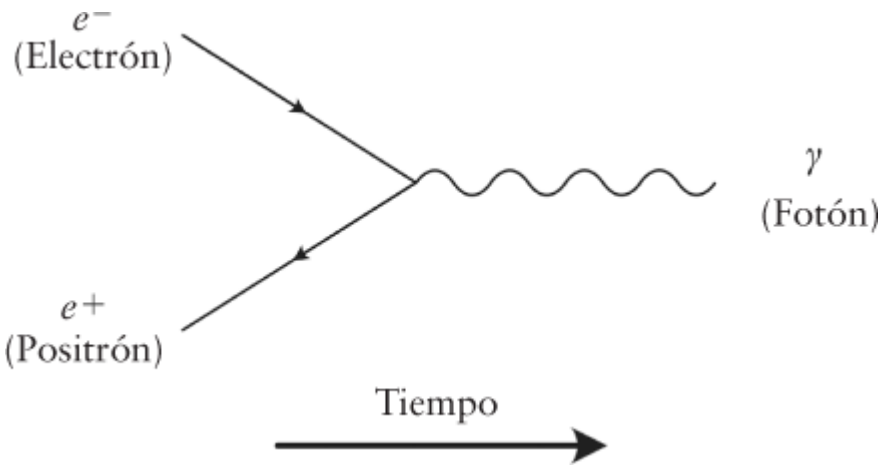
— Comience con un conjunto básico de interacciones permitidas entre partículas, como «un electrón y un positrón se aniquilan para formar un fotón».
— Dibuje un diagrama con partículas que se encuentran en un punto, llamado **vértice**, para representar cada interacción básica.
— Especifique un proceso de dispersión de interés, como «un electrón y un positrón entrantes se dispersan en un electrón y un positrón salientes». Lo importante es fijar lo que entra y lo que sale.

— Combine las interacciones básicas *de todas las formas posibles* que puedan contribuir al proceso especificado. Cada forma corresponderá a un **diagrama de Feynman** concreto.
— A cada diagrama, asocie un número complejo utilizando un conjunto especificado de reglas.
— Sume las contribuciones de cada diagrama (en general, un número infinito de ellos). El resultado es la **amplitud de dispersión** desde el estado inicial especificado hasta el estado final.
— Eleve al cuadrado la amplitud para obtener la probabilidad de dispersión deseada.

Los vértices de interacción que representan un conjunto básico de interacciones permitidas entre partículas forman los bloques de construcción de los diagramas de Feynman. Las partículas se representan mediante líneas, a menudo con características que nos ayudan a identificar de qué tipo de partícula estamos hablando, aunque hay más tipos de partículas que tipos de líneas, así que a veces simplemente las etiquetamos o se supone que las conocemos por experiencia. Cuando varias líneas se cruzan en un punto, eso es un vértice. Podemos averiguar qué vértices hay observando la dinámica subyacente de la teoría (por lo general, el lagrangiano). A cada vértice se le asigna un valor numérico, en función de la constante de acoplamiento de la interacción en cuestión. Luego los ensamblamos en varias combinaciones para hacer diagramas de Feynman completos.

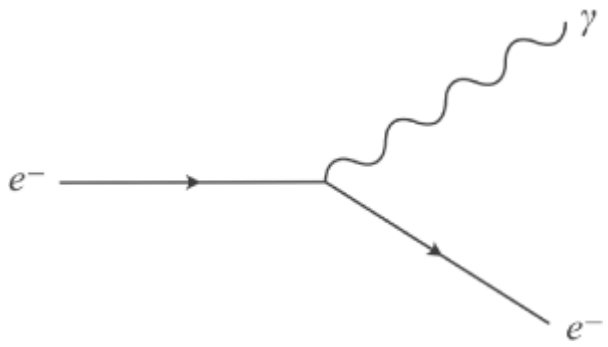
ANTIPARTÍCULAS E INVERSIÓN TEMPORAL

Un ejemplo paradigmático de un vértice de diagrama de Feynman es el correspondiente a la unión de un electrón y un positrón para convertirse en un fotón. Podemos pensar en esto como un proceso que se desarrolla en el tiempo, de izquierda a derecha en el diagrama.

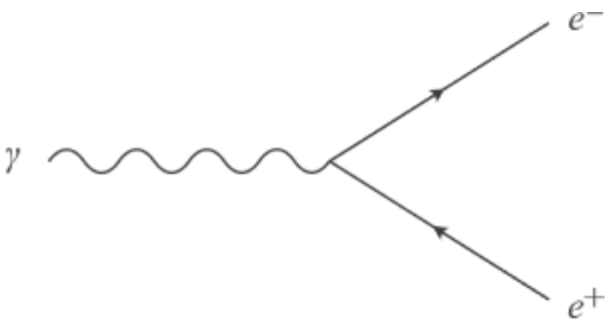


Ya vemos una serie de rasgos característicos de los diagramas de Feynman. Conforme a lo prometido, las diferentes partículas están representadas por diferentes tipos de líneas: onduladas para los fotones, y rectas para los electrones y los positrones. (El símbolo de un fotón es γ, la letra griega gamma). Las pequeñas flechas de las líneas del electrón y del positrón son cruciales: representan el flujo de los elementos con calidad de partícula, por oposición a los que tienen calidad de antipartícula. Así, vemos que el electrón tiene una flecha que apunta de izquierda a derecha, en la misma dirección en la que fluye el tiempo en nuestros diagramas, mientras que la flecha del positrón apunta en la otra dirección porque es una antipartícula. Esto no significa que el positrón se mueva en otra dirección; las flechas de las líneas de los electrones y positrones no tienen absolutamente nada que ver con la dirección de su movimiento. Solo nos recuerdan que el electrón es «materia» y el positrón es «antimateria». Por supuesto, esto es una convención: no es difícil imaginar un universo alternativo con personas hechas de antiprotones, antineutrones y positrones, y esas antipersonas naturalmente etiquetarían a los positrones como «materia» y a los electrones como «antimateria». Las etiquetas son arbitrarias. También son completamente opcionales. Si desea dibujar diagramas de Feynman en los que las líneas de electrones y positrones no tengan ninguna flecha, nadie puede impedírselo.

Esto nos lleva a otra característica fundamental de los diagramas de Feynman: dado cualquier vértice de interacción, podemos construir otros vértices (que serán numéricamente iguales una vez que utilicemos estos diagramas para calcular amplitudes) (1) cambiando la dirección en la que se mueve la partícula y (2) intercambiando la naturaleza de partícula por la de antipartícula. Así, por ejemplo, el vértice anterior implica la existencia de otro vértice en el que movemos el positrón de entrada a salida, convirtiéndolo en un electrón en el proceso. Esto representa la emisión de un fotón por un electrón en movimiento.



O podríamos trasladar tanto el electrón y el positrón entrantes al lado saliente, así como el fotón al lado entrante. El electrón y el positrón se convierten en positrón y electrón, pero es la misma colección de partículas. El fotón puede considerarse como su propia antipartícula, por lo que no se ve afectado. Ahora el diagrama ilustra la división de un fotón en un par electrón/positrón.



También hay otras versiones, como la de un positrón entrante que emite un fotón, con las que le invito a probar.

Esta característica de los diagramas de Feynman —siempre que hay un vértice, existe un vértice igual que se obtiene intercambiando entrante por saliente y materia por antimateria— ha llevado al lema de que «las antipartículas son solamente partículas que viajan hacia atrás en el tiempo». Eso no es del todo erróneo, pero como ocurre con muchas traducciones de las matemáticas al lenguaje cotidiano, puede dar una impresión equivocada. Nada «viaja» en el tiempo, ni hacia delante ni hacia atrás; las cosas existen en cada instante de tiempo, y las leyes de la física garantizan que hay cierta persistencia de las cosas de un instante a otro. Desde luego, cualquier insinuación de que podríamos utilizar antipartículas para enviar señales hacia atrás en el tiempo, o afectar de algún modo al pasado, está completamente fuera de lugar.

Lo que puede afirmarse de forma irrefutable es que, en los diagramas de Feynman, la descripción matemática de una partícula que evoluciona en una dirección del tiempo es idéntica a la de una antipartícula que evoluciona en la dirección opuesta. Todo lo que realmente nos importa es el hecho de que, cuando se invierte la línea de una partícula de entrante a saliente, la dirección de su flecha se invierte con ella.

UNA SERIE INFINITA

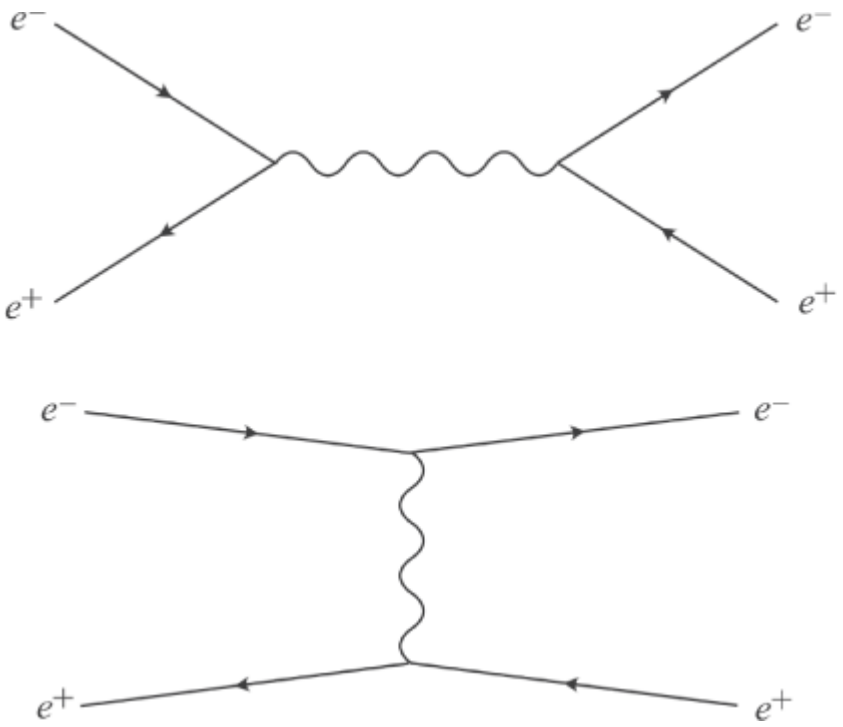
Recapitulemos todo lo anterior considerando más concretamente lo que haríamos para calcular la probabilidad de que un electrón y un positrón se dispersen entre sí. Y para mantener las cosas bajo control, consideraremos solo la fuerza electromagnética; no nos preocuparemos de la gravedad ni de los bosones de Higgs ni de nada más exótico. Nuestro contexto es, por tanto, la **electrodinámica**

cuántica, o EDC, que fue el foco principal del trabajo de la década de 1940 sobre la dispersión en la teoría cuántica de campos.

Podría pensar: «Bueno, los electrones y los positrones son partículas cargadas; sin duda se dispersarán, así que la probabilidad es 1». Estamos siendo más específicos que eso. Nuestro estado entrante especifica no solamente cuáles son las partículas entrantes, sino también sus momentos individuales, y lo mismo ocurre con las partículas salientes. En general, habrá una probabilidad de que las partículas salientes se muevan en varias direcciones, y podemos calcularla.

Así que queremos construir todos los diagramas posibles que contribuyan a este proceso, calcular la contribución numérica de cada uno y sumarlos todos.

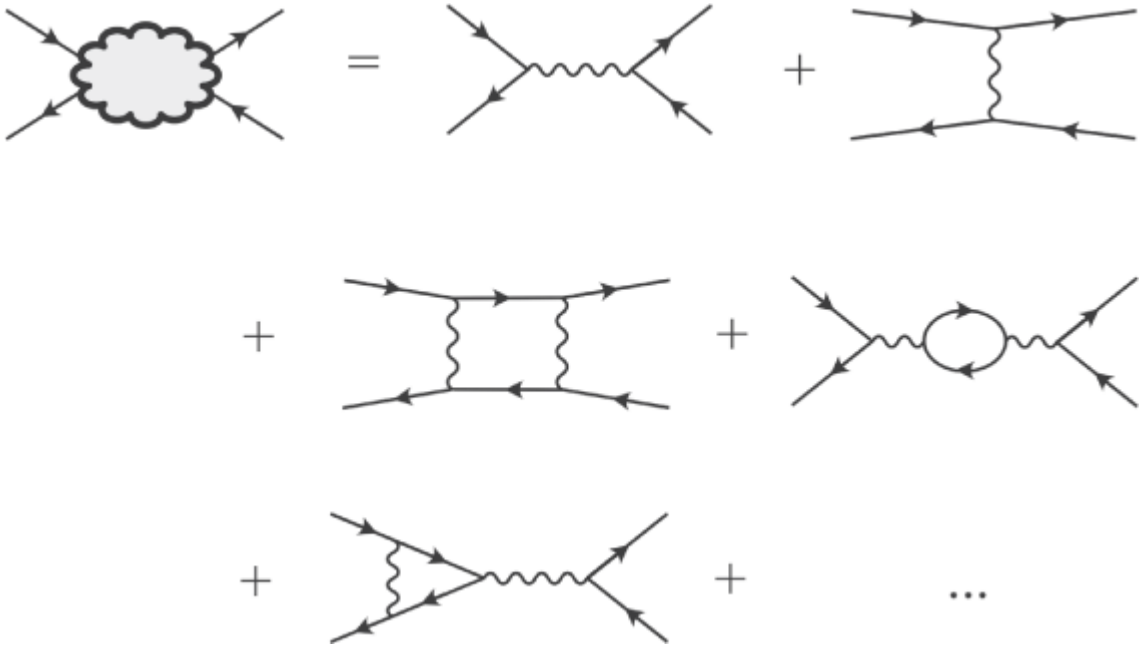
Hay una contribución bastante obvia a este proceso: el electrón y el positrón se convierten en un fotón mediante la interacción que ya hemos visto, y luego ese fotón se convierte de nuevo en un par electrón/positrón. Pero hay otra en la que podemos pensar y que es igual de sencilla: el electrón no se aniquila con el positrón, sino que simplemente le lanza un fotón.



Quizá le preocupe que no quede claro si el fotón fue emitido por el electrón y absorbido por el positrón, o viceversa. Quédese tranquilo; la receta dice que debemos agregar todo lo que pueda ocurrir, así que este diagrama pretende abarcar ambas posibilidades. Como el fotón es su propia antipartícula, no necesitamos dibujar diagramas separados para las diferentes orientaciones.

De hecho, no hay que tomar los diagramas demasiado al pie de la letra, como si fueran representaciones de partículas clásicas puntuales que se movieran a lo largo de esas trayectorias. Cuando dibujamos diagramas de Feynman, las líneas de las partículas no representan en realidad posiciones concretas en el espacio, sino estados de partículas con un momento definido. Y tales estados están completamente deslocalizados en el espacio, según el principio de incertidumbre. Eso está bien, ya que los diagramas son solamente instrumentos para ayudarnos a hacer cálculos en la teoría cuántica de campos, no fotografías de procesos físicos. Podemos pensar que el resultado representa la probabilidad de que este proceso ocurra si las partículas están justo una encima de la otra; siempre podemos corregir las cosas

más tarde para tener en cuenta las partículas que pasan con un cierto grado de separación.



Y, por supuesto, podemos añadir diagramas más complicados, siempre que nos atengamos a nuestro vértice básico. Así, por ejemplo, nunca tendremos dos electrones que se conviertan en un fotón, o un par electrón/positrón que se convierta en un solo electrón: no existen tales vértices. Caben infinitas posibilidades, y nuestra tarea es sumarlas todas.

Tómese la libertad de dibujar todos los diagramas que se le ocurran. Hay infinidad de ellos, no se le agotarán.

CÁLCULO DE AMPLITUDES

El objetivo de estos diagramas no es solamente ofrecer visualizaciones sugerentes de lo que puede ocurrir cuando las partículas interaccionan entre sí. Son la base de predicciones numéricas precisas de las probabilidades de dispersión. Aquí es donde el formalismo matemático completo se complica, y un típico estudiante de posgrado

en física pasa más o menos un año de su formación tratando de entenderlo. En nuestro afán por centrarnos en las ideas subyacentes importantes, vamos a pasar por alto las ecuaciones equivalentes a todo un libro de texto para presentar la esencia del proceso.

Esta esencia es la siguiente: cada vértice de interacción se asocia a un número, la **constante de acoplamiento** correspondiente a esa interacción. La contribución de cualquier diagrama concreto se calcula en dos pasos:

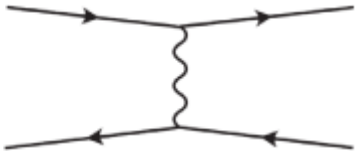
1. Multiplique las constantes de acoplamiento asociadas a cada vértice de interacción que aparece en el diagrama.
2. Sume las contribuciones de todas las energías y momentos posibles de las partículas que aparecen en el interior del diagrama.

El segundo paso es el que conduce a los infinitos y a otros problemas conceptuales; lo estudiaremos con detalle en el próximo capítulo. Por ahora, concentrémonos en el primer paso.

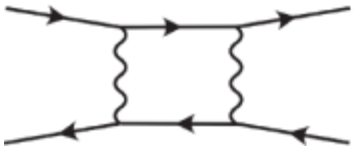
En EDC, la constante de acoplamiento asociada al vértice fundamental es $\sqrt{\alpha}$, donde α (letra griega alfa) es la constante de estructura fina, un número adimensional (es decir, sin unidades como la longitud o el tiempo) que experimentalmente resulta de medir

$$\alpha \approx \frac{1}{137}. \tag{5.1}$$

La razón por la que elegimos una convención tal que la raíz cuadrada de α, y no α en sí, se asocia con el vértice es que los procesos del mundo real, como la dispersión electrón/positrón, generalmente tienen dos o más vértices en ellos. (Un electrón y un positrón no pueden aniquilarse en un único fotón real conservando la energía y el momento). Así que los diagramas más simples que contribuyen a este proceso de dispersión son proporcionales a α.

$\propto \alpha$,  $\propto \alpha$.

Diagramas más complicados, con más y más vértices, aportarán naturalmente potencias de α cada vez más altas. Decimos «proporcional a» en lugar de «igual a» porque estamos despreciando factores numéricos como 2π, así como factores que dependen de los momentos de las partículas entrantes.

 $\propto \alpha^2$, $\propto \alpha^2$.

Todos estos aspectos son cruciales para el trabajo de un físico de partículas, pero nosotros queremos aislar los principios físicos básicos.

Esta dependencia de *a* explica, por fin, por qué hay al menos una esperanza de que la suma de una serie infinita de diagramas pueda darnos una respuesta finita: se debe a que los diagramas más complicados contribuyen en cantidades mucho menores a la respuesta final. Esto sucede porque *a* es un número pequeño, mucho menor que 1. Es algo así como considerar la suma infinita

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\cdots=1. \qquad (5.2)$$

El hecho de que estemos sumando infinitos números no significa que la respuesta no pueda ser finita y bien definida, siempre que esos números sean lo suficientemente pequeños a medida que avanzamos en la secuencia. No es obvio si lo será o no: aunque las contribuciones de los diagramas más complicados son bastante pequeñas, hay muchos más diagramas que considerar, por lo que podría no estar claro si la serie «converge» a una respuesta finita. De hecho, después de mucho trabajo, es posible demostrar que la serie de diagramas de Feynman en

EDC casi converge, pero no del todo. Lo que esto significa en la práctica es que no se puede obtener una respuesta exacta de esta manera, pero los primeros términos (las menores potencias de α) dan en realidad una aproximación bastante buena a los procesos del mundo real.

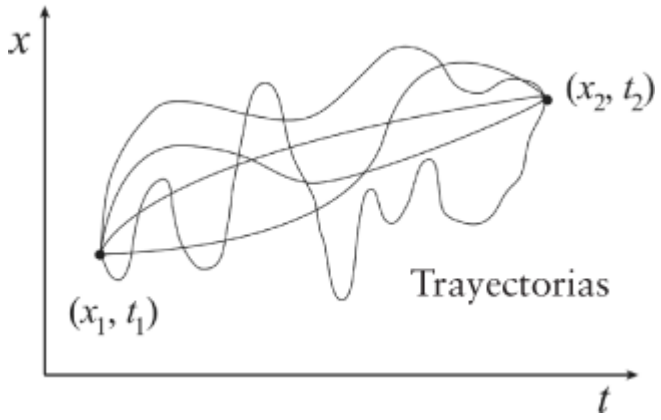
La serie de diagramas de Feynman para la EDC es un ejemplo de la **teoría de perturbaciones**, que mencionamos en *Espacio, tiempo y movimiento*. Se trata de una técnica mediante la cual resolvemos un problema difícil (campos que interaccionan en electrodinámica cuántica) resolviendo primero uno fácil (campos libres, con las interacciones desactivadas) y luego corrigiendo de manera gradual esa respuesta en términos de potencias de una pequeña perturbación. En EDC, es el hecho de que la constante de estructura fina sea tan pequeña lo que hace que la teoría de perturbaciones sea tan poderosa. A veces, como en las interacciones de quarks y gluones dentro de un protón, las constantes de acoplamiento relevantes no son pequeñas en absoluto, y tenemos que utilizar técnicas diferentes. Pero es una característica afortunada de nuestro universo físico que muchas interacciones puedan considerarse como pequeñas perturbaciones de la dinámica no interactiva del sistema que estamos considerando.

LAGRANGIANOS PARA CAMPOS

Los diagramas de Feynman proporcionan una forma elegante de calcular la probabilidad de los sucesos de dispersión de partículas en el marco de la teoría cuántica de campos. Pero hasta ahora hemos estado hablando de los diagramas en sí mismos, sin hacer mucha conexión con la teoría de campos subyacente. Esa conexión proviene de la forma precisa en que los valores numéricos, como la constante de estructura fina, se adjuntan a los diagramas. Y eso se entiende mejor pensando en el enfoque lagrangiano de la teoría de campos. La conexión resultará ser bastante simple y elegante: cada término del

lagrangiano está directamente asociado con parte de un diagrama, ya sean las líneas que representan partículas en movimiento o los vértices de interacción donde se encuentran unas con otras.

En *Espacio, tiempo y movimiento* definimos el **lagrangiano** en mecánica clásica como la energía cinética menos la energía potencial, $L = K - V$, y la **acción** de una trayectoria como la integral del lagrangiano a lo largo del tiempo, $S = \int L\, dt$. Considerando todas las trayectorias posibles (o «historia») de un sistema entre un punto inicial y un punto final, la que realmente sucede —la trayectoria que obedece las leyes de la física clásica— es la que tiene la mínima acción de todas ellas. Para una partícula, esa trayectoria es la posición en función del tiempo, $x(t)$, mientras que para un campo es todo un perfil espacio-temporal, $\phi(x, t)$, pero la esencia de la idea es la misma en ambos casos.



La versión cuántica es un poco diferente, basada en la **integral de trayectorias** o «suma extendida a todas las historias» de Feynman. Ahora cuentan todas las trayectorias, en lugar de una única que tenga la acción mínima S. Más que buscar una trayectoria especial con el mínimo S, calculamos el número complejo e^{iS} a lo largo de cada trayectoria, y luego lo integramos para todas las trayectorias a fin de obtener la amplitud cuántica para empezar en una configuración y terminar en otra.

Esto funciona gracias a una propiedad que posiblemente recordemos del cálculo: en el mínimo de una función, la derivada es 0. Esto significa que los valores de la función cerca del mínimo son casi iguales entre sí. De este modo, en las proximidades de la trayectoria de mínima acción, los valores de e^{iS} son aproximadamente iguales; al integrar esa cantidad para todas las trayectorias, se suman y se refuerzan mutuamente, lo que nos da una gran probabilidad. Fuera de la trayectoria de mínima acción, las trayectorias cercanas dan valores muy diferentes para e^{iS}, positivos y negativos, así como positivos imaginarios y negativos imaginarios, por lo que interfieren y se anulan mutuamente. Así es como, desde la perspectiva de la integral de la trayectoria, el comportamiento de aspecto clásico es probable, pero no inevitable.

Cuando hablamos de la energía de un campo en el capítulo anterior, observamos que la localidad implica una forma especial de calcular esa energía: existe una densidad de energía $\rho(x)$, construida a partir del campo y sus derivadas, y la integramos en el espacio para obtener la energía total, $E = \int \rho(x)\, dx$. Lo mismo va a suceder para el lagrangiano: construiremos una **densidad de Lagrange** $\mathcal{L}(x)$, luego integraremos en el espacio para obtener el lagrangiano real,

$$L = \int \mathcal{L}(x) dx. \tag{5.3}$$

Como de costumbre, escribimos una única coordenada x, pero se debe considerar que representa las tres coordenadas espaciales. La acción sigue siendo la integral en el tiempo del lagrangiano, por lo que es la integral en todo el espacio-tiempo de la densidad de Lagrange: $S = \int L\, dt = \int \mathcal{L}\, dx\, dt$.

Cuando los físicos de partículas hablan de una teoría o un modelo específicos, normalmente están pensando en definirlos escribiendo la densidad de Lagrange. Muchas de las horas de trabajo en las vidas de los físicos teóricos de partículas del último medio siglo se han empleado en pensar acerca de las densidades de Lagrange. Tanto es así

que solemos ser perezosos y nos limitamos a decir «el lagrangiano» cuando en realidad queremos decir «la densidad de Lagrange», asumiendo que usted es lo suficientemente sofisticado como para saber que tiene que integrarla en el espacio para obtener el lagrangiano real (y en el tiempo para luego obtener la acción).

Se supone que un lagrangiano es la energía cinética menos la energía potencial. Para los campos, tenemos una nueva contribución: la energía de gradiente. Afortunadamente, el requisito de que mantengamos la compatibilidad con la relatividad (el espacio y el tiempo son dos aspectos diferentes del espacio-tiempo subyacente) obliga a que exista una relación definida entre la energía cinética (derivada temporal) y la energía de gradiente (derivada espacial). En consecuencia, la energía de gradiente aparece con un signo menos en comparación con la energía cinética. (Es el mismo signo menos que aparece en el tensor métrico del espacio de Minkowski). Así que, en este punto, tenemos:

$$\begin{pmatrix}\text{Lagrangiano}\\ \text{del campo}\end{pmatrix} = \begin{pmatrix}\text{Energía}\\ \text{cinética}\end{pmatrix} - \begin{pmatrix}\text{Energía de}\\ \text{gradiente}\end{pmatrix} - \begin{pmatrix}\text{Energía}\\ \text{potencial}\end{pmatrix}$$

Por ejemplo, en la teoría de campo libre escalar que analizamos en el capítulo anterior, el lagrangiano es

$$\mathcal{L} = \frac{1}{2}\left(\partial_t \phi\right)^2 - \frac{1}{2}\left(\partial_x \phi\right)^2 - \frac{1}{2} m_\phi^2 \phi^2 \,. \qquad (5.4)$$

Para esta teoría de campo libre, podríamos resolver todas las ecuaciones exactamente. Los modos de longitud de onda definida actúan como osciladores armónicos simples, y podemos interpretar las excitaciones del campo como partículas que se mueven con total libertad. Si tenemos varios campos libres, podemos simplemente sumar sus lagrangianos. El resultado describe varios tipos de partículas, pero estas siguen desplazándose sin interaccionar.

Ahora queremos introducir las interacciones. Ahí es donde van a ocurrir todas las cosas interesantes. Suponemos que podemos resolver exactamente la teoría sin interacciones, y luego incluir las interacciones y tratarlas utilizando la teoría de perturbaciones, generando una secuencia infinita de diagramas de Feynman. Por tanto, una descomposición más útil es:

$$\begin{pmatrix}\text{Lagrangiano}\\ \text{del campo}\end{pmatrix} = \begin{pmatrix}\text{Lagrangiano}\\ \text{libre}\end{pmatrix} + \begin{pmatrix}\text{Lagrangiano}\\ \text{de interacción}\end{pmatrix}$$

Casi siempre —no absolutamente siempre, pero lo suficiente como para que no nos preocupemos aquí de las excepciones— las interacciones se definirán mediante adiciones a la energía potencial, no a la energía cinética/de gradiente. Obsérvese que en todas las densidades de energía de campo libre —la energía cinética $\frac{1}{2}(\partial_t\phi)^2$, la energía de gradiente $\frac{1}{2}(\partial_x\phi)^2$ y la energía potencial $\frac{1}{2}m^2\phi^2$— aparece dos veces la variable de campo ϕ. La energía cinética lleva la derivada temporal del campo al cuadrado, la energía de gradiente lleva la derivada espacial del campo al cuadrado, y la energía potencial lleva el propio campo al cuadrado. Las interacciones vienen dadas por la multiplicación de tres o más variables de campo. He aquí, por ejemplo, un posible lagrangiano que describe dos campos escalares que interaccionan, ϕ y θ:

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}(\partial_t\phi)^2 - \frac{1}{2}(\partial_x\phi)^2 - \frac{1}{2}m_\phi^2\phi^2\\ &+ \frac{1}{2}(\partial_t\theta)^2 - \frac{1}{2}(\partial_x\theta)^2 - \frac{1}{2}m_\theta^2\theta^2\\ &- A\phi^2\theta - B\phi^2\theta^2 . \end{aligned} \tag{5.5}$$

Los tres primeros términos del lado derecho (la primera línea) representan el lagrangiano libre para ϕ, donde m_ϕ representa su masa.

Los tres términos siguientes (segunda línea) son el lagrangiano libre para θ, donde m_θ representa su masa, que podría muy bien ser diferente de la masa de ϕ. Todos esos términos incluyen dos apariciones del campo apropiado. Los dos últimos términos (última línea) son las interacciones, donde A y B desempeñan el papel de constantes de acoplamiento. El término A tiene tres campos en total (ϕ dos veces, θ una vez) y el término B tiene cuatro campos en total (ϕ dos veces y θ dos veces).

Ahora solamente tenemos que ver cómo esas interacciones se convierten en vértices de los diagramas.

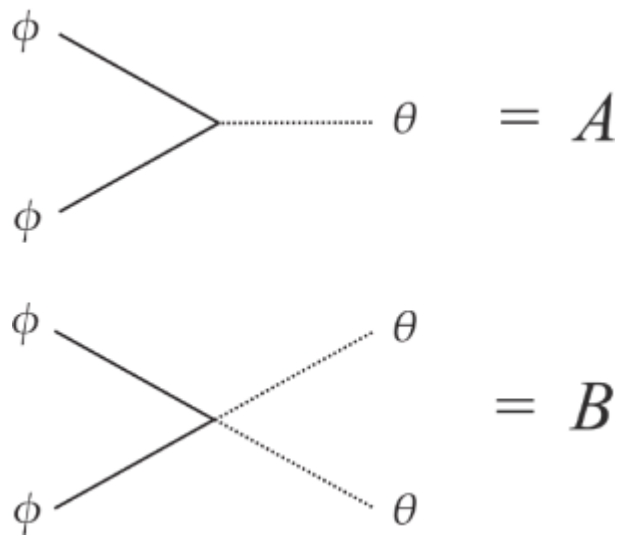
DE LOS LAGRANGIANOS A LOS DIAGRAMAS

Piense en un único campo libre ϕ. ¿Qué aspecto tendrían sus diagramas de Feynman? No hay interacciones y, por tanto, no hay vértices. La partícula simplemente se propaga de un lugar a otro, sin ser perturbada. Pero podríamos, si en verdad quisiéramos, representar esto en forma de diagrama. Sería simplemente una única línea recta, representando una partícula ϕ que se propaga.

No es muy emocionante, la verdad. Para campos más complicados, que incluyan electrones y fotones, este **propagador** tendrá en realidad una expresión matemática no trivial. De una forma u otra, solo representa la idea de que la partícula viaja sin obstáculos de un lugar a otro.

Pero fíjese en lo siguiente sobre el diagrama del propagador. Para un campo libre, todos los términos de su lagrangiano tienen dos apariciones del campo. Y el propagador tiene dos extremos, uno entrando y otro saliendo, ambos etiquetados con ese campo.

Hagamos una generalización audaz: cada término de la interacción lagrangiana se asocia a un vértice con una línea por cada campo que aparece en el término, y el valor numérico asociado a él es simplemente su constante de acoplamiento. Para el modelo de dos campos (5.4), por ejemplo, existen dos términos de interacción, $A\phi^2\theta$ y $B\phi^2\theta^2$. Denote-mos ϕ mediante una línea continua y θ mediante una línea discontinua. Entonces tenemos dos vértices:



Esto es básicamente lo que necesita saber para construir todos los diagramas de Feynman de esta pequeña teoría. Tómese la libertad de probar dibujando algunos complicados. Tenga en cuenta que solo puede hacer lo que le permitan los vértices. Por ejemplo, una sola ϕ nunca se transforma espontáneamente en una sola θ; tampoco dos θ se aniquilan produciendo una sola ϕ. Pero son posibles muchas otras interacciones. (En esta sencilla teoría no hay distinción entre partículas y antipartículas, ya que las partículas no llevan ninguna cantidad conservada que pueda ser positiva o negativa).

Una teoría realista como la EDC será más complicada, incluso con un número reducido de campos, porque los propios campos serán objetos geométricos más complicados que simples números. Pero la idea básica es la misma. Si etiquetamos el campo del electrón como e^-,

el campo del positrón como e^+ y el campo electromagnético (fotón) como γ, el lagrangiano de la interacción EDC es en esencia

$$\mathcal{L}_{EDC} = -\sqrt{\alpha} e^- e^+ \gamma. \tag{5.6}$$

Este término se traduce directamente en el vértice fundamental de la EDC, donde un electrón y un positrón se convierten en un fotón (o en variaciones de este).

Ya ve por qué los físicos de partículas se pasan todo el tiempo pensando sobre los lagrangianos. El lagrangiano define la dinámica de los campos individuales, así como la forma en que los distintos campos interaccionan entre sí. Una vez que se sabe eso, el resto consiste solo en girar una manivela para hacer algunas predicciones. Los lagrangianos implican propagadores y vértices, que se juntan para formar diagramas, que se suman para calcular amplitudes, que se elevan al cuadrado para obtener probabilidades.

PARTÍCULAS VIRTUALES

Vale, no es tan sencillo como podría dar a entender el párrafo anterior. Hay sutilezas en los diagramas de Feynman y en cómo se utilizan. Algunas de esas sutilezas son simples y divertidas; otras son desafiantes y profundas. Veamos las primeras, dejando las segundas para el próximo capítulo.

Cada diagrama de Feynman cuenta una historia. Entran unas partículas, interaccionan intercambiándose con otras y sale otro conjunto de partículas. Pero no todas estas partículas se crean igual. Todas las líneas de entrada y salida representan partículas «reales», que existen de verdad. Sin embargo, una línea que es completamente interna a un diagrama —que comienza en algún vértice y termina en otro, sin llegar nunca al mundo exterior— representa una **partícula virtual**. Las partículas virtuales no son partículas reales. Representan

un proceso real —las vibraciones de la interacción de un conjunto de campos cuánticos—, pero no partículas existentes, del mundo real, que pudiéramos ver en un experimento.

La distinción entre partículas reales y virtuales es relevante debido a la conservación del momento. En *Espacio, tiempo y movimiento* vimos que, en la relatividad, la energía y el momento se unifican en un único «cuadrivector de momento», siendo la energía la componente cero-ésima (similar al tiempo) del vector. Para las partículas reales, sin embargo, las componentes del vector momento no son en absoluto independientes; la energía E se relaciona con el momento espacial $\vec{p}$ y la masa m mediante

$$E^2 = \vec{p}^2 + m^2 . \qquad (5.7)$$

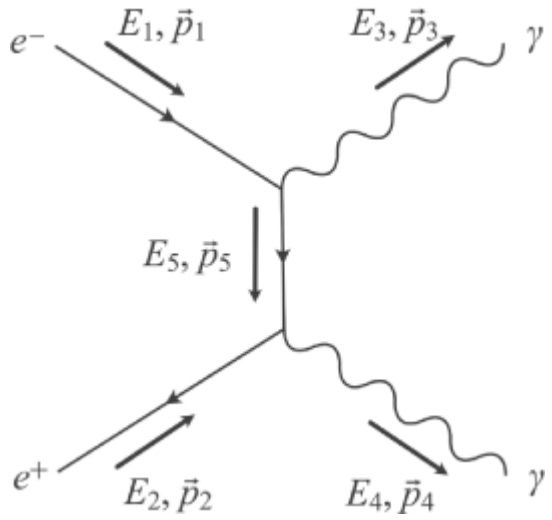
Cuando escribimos diagramas de Feynman, pensamos en estados de partículas con momentos bien definidos. Y hay una regla que dice que *todas las componentes del momento se conservan exactamente en cada vértice de un diagrama*. La dinámica de dispersión no viola la conservación de la energía y el momento.

Esto tiene algunas consecuencias inmediatas. Por ejemplo, el vértice fundamental de la EDC que consideramos, en el que un electrón y un positrón se convierten en un único fotón, *no* es un diagrama de Feynman admisible por sí mismo. Esto se debe a que no hay forma de sumar el momento de dos partículas masivas y más lentas que la luz (un electrón y un positrón) y obtener un resultado igual al de una única partícula sin masa y con la velocidad de la luz (el fotón). Para ver esto, observe que siempre hay un marco de referencia donde el momento espacial del electrón es igual y opuesto al momento del positrón, por lo que el momento espacial total es 0. Pero un fotón no puede tener momento espacial cero; eso implicaría que estaría quieto, no moviéndose a la velocidad de la luz.

Así se explica por qué, cuando los electrones y los positrones se aproximan y se aniquilan, tienen que producir dos fotones, no

solamente uno. Aunque nuestro vértice fundamental de la EDC no sea un respetable diagrama de Feynman por sí solo, no hay ningún problema en que forme parte de un diagrama mayor.

Veamos un proceso en el que un electrón y un positrón se aniquilan produciendo dos fotones. A diferencia de nuestros diagramas anteriores, aquí hemos ilustrado explícitamente el flujo de energía y momento. Hacerlo es completamente opcional, dependiendo de sus propósitos. El electrón entrante es 1, el positrón es 2, los fotones salientes son 3 y 4, y la partícula virtual intermedia es 5. Las flechas de momento no tienen nada que ver con las flechas de las líneas que denotan el flujo de los elementos con calidad de partícula; son conceptos completamente diferentes. Puede que le preocupe que sea ambiguo que la partícula virtual sea un electrón o un positrón. La buena noticia es que, en el caso de las partículas virtuales, no importa, ya que en última instancia queremos incluir ambas posibilidades.



El hecho de que tanto la energía como el momento se conserven en cada vértice nos da las siguientes relaciones:

$$\vec{p}_1 = \vec{p}_3 + \vec{p}_5$$
$$E_1 = E_3 + E_5$$
$$\vec{p}_2 + \vec{p}_5 = \vec{p}_4$$
$$E_2 + E_5 = E_4. \qquad (5.8)$$

Las flechas de momento del diagrama le indican cómo se suman las cosas: el momento 1 se divide en los momentos 3 y 5, mientras que los momentos 2 y 5 se combinan para dar el momento 4.

Sumando y restando estas ecuaciones en combinaciones adecuadas, puede convencerse de que la energía E_5 y el momento $\vec{p}_5$ de la partícula virtual están completamente determinados por las energías entrantes y salientes y el momento; no hay margen de maniobra. Pero, en general, *no* satisfarán la relación clásica entre energía, momento y masa:

$$E_5^2 \neq \vec{p}_5^{\,2} + m^2 \,. \qquad (5.9)$$

De hecho, la energía simplemente será la que tenga que ser para satisfacer las relaciones de (5.8); incluso puede ser negativa. Esto es importante para explicar cómo, por ejemplo, un electrón y un positrón pueden atraerse «lanzando» un fotón de uno a otro. ¿Lanzar un fotón no debería producir un alejamiento? No si el fotón tiene energía negativa. En ese caso, en realidad se produce un empuje en la misma dirección hacia la que se lanza el fotón.

Y no pasa nada, porque la partícula es virtual, no real. No se puede crear una cierta cantidad de energía negativa capturando partículas virtuales; solamente existen como herramientas matemáticas dentro de los diagramas de Feynman. Pero proporcionan una forma muy cómoda de pensar sobre los procesos en la teoría cuántica de campos, y a veces existe la tentación de olvidar que no son lo que realmente está sucediendo.

Y otras veces, pensar mucho en las partículas virtuales nos lleva a situaciones asombrosas. Ese es el tema del próximo capítulo.

SEIS

TEORÍA DE CAMPO EFECTIVO

La física siempre ha tenido una relación incómoda con el infinito. Por un lado, aparece por todas partes, a menudo de forma extremadamente útil. Cuando Newton y Leibniz desarrollaron el cálculo, parte del reto consistía en tratar el infinito de forma sistemática. Y el infinito no solo se refiere a cosas que son «infinitamente grandes»; hay un número infinito de números reales entre 0 y 1. De hecho, cualquier estructura matemática que sea uniforme y continua, desde el espacio-tiempo hasta el espacio de Hilbert, tiene un número infinito de elementos, y todas nuestras mejores ideas actuales sobre la física fundamental utilizan este tipo de estructuras. Y, aunque en realidad no lo sepamos, es del todo posible que nuestro universo real se extienda infinitamente lejos en el espacio, o en el tiempo, o en ambos.

Por otra parte, nos consideramos seres finitos, restringidos a una parte finita del universo (tanto en el tiempo como en el espacio), con una capacidad finita de observar y experimentar. Por tanto, si escribimos una teoría y descubrimos que predice que alguna cantidad física perceptible crecerá infinitamente justo delante de nosotros, tendemos a preocuparnos. Según la relatividad general, por ejemplo, el interior de los agujeros negros contendrá singularidades en las que la curvatura del espacio-tiempo parece acercarse al infinito. La mayoría de los físicos lo consideran una prueba de que la relatividad general, que al fin y al cabo es una teoría clásica, no es capaz de describir lo que ocurre en estas circunstancias extremas. En la mecánica estadística clásica, un cálculo sencillo de la radiación del cuerpo negro parece predecir una cantidad infinita de radiación de longitud de onda corta:

la catástrofe ultravioleta. Pensar en solucionar ese problema nos puso en el camino de la mecánica cuántica.

Uno de los retos inmediatos a los que se enfrentaron los pioneros de la teoría cuántica de campos fue el reconocimiento de que ciertos términos de la serie infinita de diagramas de Feynman, que se suponía que se harían más pequeños a medida que los diagramas se complicaran, en realidad daban respuestas infinitas. La resolución de este problema condujo al estudio de la **renormalización** y cambió radicalmente nuestra forma de pensar sobre la teoría cuántica de campos. Este esfuerzo culminó en la idea de las **teorías de campo efectivo**, y este capítulo explicará lo que eso significa. La idea básica es, en resumen, la siguiente:

— En la teoría cuántica de campos estándar, cuando sumamos las contribuciones de partículas virtuales con energías arbitrariamente altas, ciertos diagramas dan resultados infinitos no físicos, que deben restarse de algún modo.
— En el paradigma de la teoría de campo efectivo, nos ocupamos de esto simplemente no incluyendo contribuciones de partículas de alta energía. Introducimos una «energía de corte ultravioleta» Λ, e incluimos solamente contribuciones de partículas virtuales por debajo de esa energía. Esto nos da una «teoría infrarroja» viable, que solo se refiere a partículas con energías inferiores a la de corte.
— La teoría de campo efectivo resultante se define mediante un conjunto de constantes de acoplamiento efectivas. Estas «constantes» no son realmente constantes: dependen del valor del corte Λ. Y lo hacen de manera que ninguna de las predicciones físicas finales dependa del valor que hayamos elegido.
— Todo en la teoría de campo efectivo resultante se comporta de forma bien definida y es finito, siempre que limitemos nuestra

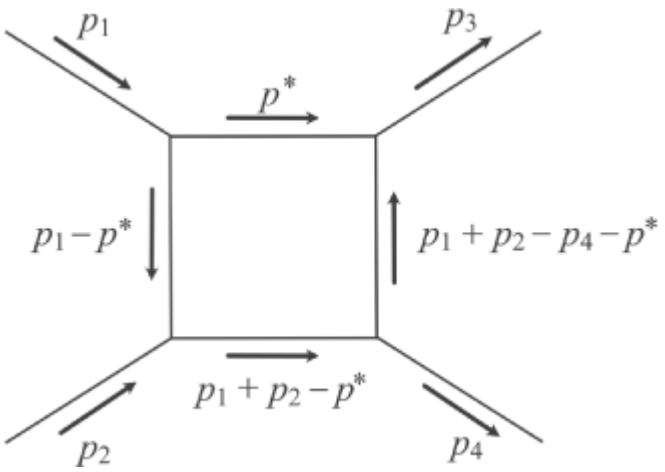
atención a partículas y procesos por debajo de la energía de corte.

El concepto de una teoría de campo efectivo con un corte en la región ultravioleta es absolutamente primordial para la física moderna. Es una de las ideas fundamentales que exploraremos en este libro.

DIAGRAMAS DE BUCLE

En primer lugar, investiguemos de dónde proceden esos valores infinitos. Al final del capítulo anterior examinamos un sencillo diagrama de Feynman para un electrón y un positrón que se aniquilaban produciendo dos fotones. Olvidemos todos los detalles de las partículas concretas y centrémonos en la idea de dos partículas que entran y dos partículas que salen, con una partícula virtual en medio. Podríamos trazar el flujo de momento a través del diagrama y observar que la energía y el momento de la partícula virtual quedaban completamente definidos por las energías y momentos de las partículas entrantes y salientes.

¿Y si nos fijamos en una situación similar, pero ahora con más de una partícula intercambiándose? Por comodidad, utilizaremos p como etiqueta «todo en uno» para designar el cuadrivector momento, que incluye la energía como su componente cero-ésimo. Cuando hablan de modo informal, los físicos suelen utilizar «momento» y «energía» como términos esencialmente intercambiables en este contexto, ya que lo importante suele ser el cuadrivector momento. Y el principio fundamental es que el momento se conserva: en cada vértice, la suma de los momentos entrantes debe ser igual a la suma de los salientes.



Mirando el diagrama se puede averiguar lo que está pasando. Leyendo de izquierda a derecha, tenemos p_1 y p_2 como momentos entrantes en el diagrama, y p_3 y p_4 como salientes. En cada vértice, la suma de los momentos entrantes es igual a la suma de los salientes. La dirección en la que dibujamos esas flechas de momento es arbitraria; podríamos invertirlas siempre y cuando también multipliquemos los momentos correspondientes por –1 al sumarlos.

En el primer vértice de la parte superior izquierda, el momento p_1 se divide en algún valor p^* que pasa por la partícula virtual superior, y $p_1 - p^*$ que va verticalmente hacia abajo. Siguiendo alrededor del cuadrado, en la parte superior derecha tenemos

$$(p^*) + (p_1 + p_2 - p_4 - p^*) = p_3. \tag{6.1}$$

Los p^* se anulan entre sí, y podemos mover p_4 del lado izquierdo al derecho para obtener

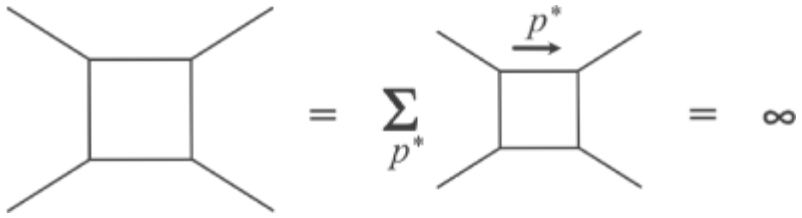
$$p_1 + p_2 = p_3 + p_4. \tag{6.2}$$

Esta ecuación se cumple automáticamente, con independencia de cuáles sean los momentos individuales. Se trata simplemente de la conservación del momento para el proceso de dispersión en su

conjunto, igualando la suma de los momentos entrantes a la suma de los salientes.

Esto significa que el valor de p^*, el momento inicialmente no especificado que atraviesa la parte superior del cuadrado, podría ser cualquiera. No tiene ninguna restricción, y la conservación global de la energía y el momento se seguirá respetando, tanto para el proceso de dispersión real en su conjunto como en cada vértice. Pensamos en p^* como el «momento interno» del diagrama. Se observa que aparece en el momento de cada línea interna.

¿Qué hacemos al respecto? El procedimiento de Feynman nos dice que sumemos una contribución de cada diagrama que podamos dibujar. Como cualquier valor de p^* funcionaría en nuestro diagrama, deberíamos sumar todas las contribuciones para cada valor de p^*. En realidad eso es hacer una integral extendida a todos los valores posibles del momento interno, pero una integral es solamente una suma extendida a una variable continua, así que no distinguiremos entre ellas.



El problema con el que nos encontramos inmediatamente es que esta integral es, para muchas teorías, igual a infinito. Eso parece mal asunto. Se suponía que los diagramas más complicados darían contribuciones menores, ya que incluyen más vértices y, por tanto, más constantes de acoplamiento, pero este (y muchos otros aún más complicados) acaba siendo infinitamente grande.

No se preocupe, es un problema que tiene solución. Pero seamos claros sobre dónde surge el problema. Es una cuestión de topología del diagrama. En algunos diagramas, si empiezas en cualquier punto y los recorres, acabarás de manera inevitable en otra línea externa. Son los

llamados **diagramas de árbol**. Pero en otros diagramas, puedes encontrar uno o más bucles cerrados y dar vueltas eternamente. Son los **diagramas de bucle**.

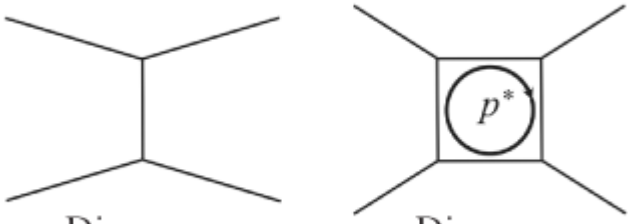


Diagrama de árbol | Diagrama de bucle

Son los diagramas de bucle los que conducen a valores infinitos. Para cada bucle cerrado habrá un momento interno (o «momento de bucle») que tendremos que integrar para todos los valores posibles, y estas integrales suelen dar infinito. Es señal de que, o bien la teoría es simplemente errónea, o bien tenemos que ser más cuidadosos a la hora de pensar en lo que está ocurriendo. (Se trata de esto último).

TEORÍA DE CAMPO EFECTIVO

El rompecabezas de los infinitos fue abordado por Tomonaga, Schwinger, Feynman, Dyson y sus colegas, y los tres primeros compartieron el Premio Nobel de Física de 1965 por sus trabajos. El procedimiento que inventaron se denominó renormalización. Se trata esencialmente de un método para restar el infinito de una manera concreta que nos permita expresar la respuesta final a un cálculo de dispersión en términos de cantidades finitas físicamente medibles.

No a todo el mundo le pareció un procedimiento completamente satisfactorio. Paul Dirac se mantuvo escéptico durante toda su vida: «La matemática sensata implica despreciar una cantidad cuando es pequeña, no despreciarla solamente porque es infinitamente grande y no la quieres». El propio Feynman lo calificó de «proceso tonto»,

«sacado de la manga» y sospechó que «no era matemáticamente legítimo»[1].

Es evidente que, no vamos a perder el tiempo preocupándonos por la legitimidad matemática de la anticuada teoría de la renormalización, porque podemos recurrir a un enfoque más moderno y bien formulado: la teoría de campo efectivo (TCE). Impulsada por el físico Ken Wilson y otros en las décadas de 1960 y 1970, el paradigma de la TCE proporciona una forma sistemática de calcular cantidades observables sin hacer aparecer primero infinitos para luego restarlos.

Wilson se inspiró en un dispositivo novedoso que empezaba a ser útil para los físicos: el ordenador digital. Imaginemos que intentamos hacer una simulación numérica de una teoría cuántica de campos. Dentro de una memoria de ordenador de tamaño finito, no podemos calcular literalmente lo que ocurre en cada localización del espacio, ya que hay un número infinito de tales localizaciones. En cambio, podemos seguir los campos de forma aproximada dividiendo el espacio en un entramado de puntos separados una cierta distancia. Wilson se dio cuenta de que podía estudiar sistemáticamente lo que ocurre a medida que hacemos que la distancia de separación sea cada vez menor. Y no es solo que estemos haciendo cálculos, tomando el límite a medida que la distancia tiende a cero. Más bien, para cualquier tamaño de entramado fijo, podemos construir una versión «efectiva» de la teoría que será tan precisa como queramos. En concreto, podemos obtener respuestas físicas sensatas sin preocuparnos nunca —o sin saber siquiera— qué ocurre cuando la distancia es exactamente cero.

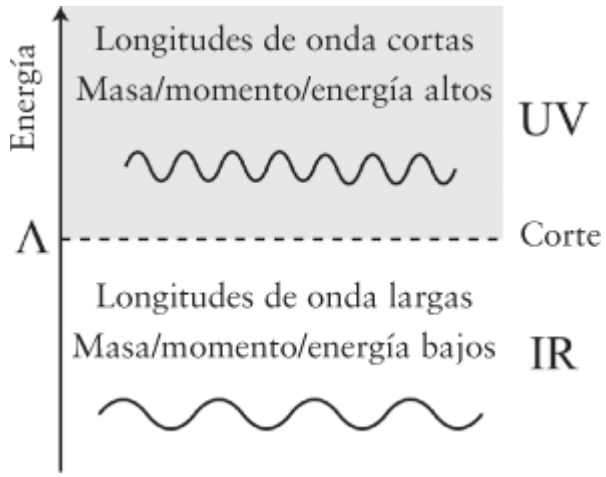
La idea que nos pone en camino hacia la teoría de campo efectivo es: sé honesto sobre lo que sabes. Reflexione detenidamente sobre por qué obteníamos respuestas infinitas en primer lugar. Era porque estábamos añadiendo contribuciones de momentos de bucle una y otra vez hasta el infinito. Pero en la teoría de campos existe una relación inversa entre energía/momento y espacio/tiempo, de modo que

cualquier momento p concreto está asociado a una distancia λ (en unidades donde $\hbar = c = 1$) según

$$p \sim \frac{1}{\lambda}. \tag{6.3}$$

Por tanto, los momentos muy grandes corresponden a distancias muy pequeñas. Cuando incluimos momentos de bucle que llegan hasta el infinito, estamos incluyendo contribuciones de partículas virtuales con energías arbitrariamente altas, a distancias arbitrariamente pequeñas. ¿Qué derecho tenemos a hacer eso? No podemos afirmar que sabemos cómo funciona la física en esas condiciones. Podría haber todo tipo de nuevas partículas y fuerzas que desconocemos. Quizás el propio espacio-tiempo deje de ser un concepto útil una vez que las distancias se vuelven increíblemente pequeñas. Deberíamos ser más francos sobre nuestra ignorancia y formular respuestas que no dependan de ninguna de esas suposiciones.

Por lo tanto, Wilson propuso ser sinceros, aunque nos incomode. No sumemos los momentos de bucle hasta energías infinitamente grandes, sino solamente hasta cierto **corte ultravioleta** seleccionado, Λ. (Recuerde que «ultravioleta» o «UV» se relaciona con gran energía/corta distancia, al igual que «infrarrojo» o «IR» lo hace con pequeña energía/larga distancia). Es decir, introducimos un límite o corte de energía Λ como nuevo parámetro de la teoría y dividimos el espacio de los momentos de las partículas en UV (por encima de Λ) e IR (por debajo de Λ). Entonces, simplemente no incluimos ninguna de las contribuciones UV cuando hacemos nuestras integrales de bucle. Recordemos que la energía está relacionada con el momento y la masa mediante $E^2 = p^2 + m^2$, por lo que las partículas UV son las que tienen una masa elevada o un momento elevado (o ambos). Las partículas virtuales con longitudes de onda corta quedan totalmente fuera de nuestra consideración, y solamente hablamos de partículas IR.



No es difícil ver que este procedimiento de fuerza bruta eliminará los infinitos. Estamos sumando un número finito de contribuciones (para cada diagrama), cada una de las cuales es a su vez finita, por lo que obtendremos una respuesta finita. Pero nos puede preocupar que el formalismo resultante no produzca más que cosas sin sentido, ya que hemos omitido muchos detalles. Sorprendentemente, la dinámica de las partículas IR, por sí solas y sin ningún rastro de partículas UV eliminadas a mano, puede ser una teoría cuántica de campos perfectamente válida, solamente que con constantes de acoplamiento diferentes a las de la teoría con la que empezamos. Por eso esas teorías se llaman «efectivas». Esto es sorprendente, porque en física (o en la ciencia en general) no se suelen suprimir ciertas partes de la descripción teórica y esperar que lo que quede sea algo sensato en absoluto. No tiene sentido pensar en términos de «mecánica clásica, pero ignorando el momento», por ejemplo.

Es una situación en la que hay buenas y malas noticias. La buena noticia es que quizá no sepamos lo que ocurre en el UV, pero no necesitamos saberlo; lo que ocurre allí podría influir en el IR, pero todos esos efectos pueden agruparse en un pequeño número de parámetros que pueden especificarse términos IR únicamente. Las partículas contarán como UV si tienen masa $m > \Lambda$, o si tienen poca masa, pero se mueven rápidamente y tienen gran momento. Así pues, la posible existencia de partículas pesadas desconocidas o de nuevos

fenómenos de longitud de onda corta (nuevas fuerzas, tal vez incluso una descomposición del propio espacio-tiempo) no nos impide describir lo que ocurre en el mundo IR de las partículas de baja masa y bajo momento.

La mala noticia es que esto hace difícil comprobar de modo experimental lo que ocurre en el mundo UV, precisamente porque no importa mucho lo que veamos en el IR. Por eso los físicos de partículas necesitan construir aceleradores gigantescos para encontrar nuevas partículas. Para averiguar qué ocurre en el UV, tenemos que visitar literalmente el régimen UV de alta energía. Nuestros mejores aceleradores actuales hacen colisionar partículas con energías del orden de 10 TeV (billones de electrón-voltios). Es extremadamente difícil para nosotros poder obtener datos experimentales directos que nos digan lo que ocurre a escalas aún mayores, pero quizá los futuros colisionadores lo consigan.

La otra propiedad destacable de la idea de Wilson es que, a pesar de que tuvimos que introducir un corte ultravioleta aparentemente arbitrario Λ, el valor de ese corte va a desaparecer completamente de nuestras predicciones físicas. La teoría IR es autónoma e independiente del corte real que utilicemos. El secreto es que las constantes de acoplamiento que definen nuestra teoría dependerán de Λ, y lo harán de tal manera que todas las predicciones seguirán siendo las mismas. Esto es lo que trata el «grupo de normalización», así que vamos a sentar las bases para ello.

UNIDADES NATURALES

La teoría de campo efectivo puede parecer demasiado buena para ser verdad. ¿Ciertas partes de su teoría (partículas virtuales de alta energía) le causan problemas? Ignórelas y cíñase a las partes que funcionan. Parece una suerte extraordinaria, como dar la espalda a la

canasta de baloncesto y lanzar un tiro por encima del hombro que acaba entrando.

La razón por la que funciona puede entenderse en términos de análisis dimensional. Recordemos que todas las magnitudes físicas, como la masa, la energía, la distancia, el tiempo y así sucesivamente, se miden en términos de determinados sistemas de unidades. El análisis dimensional no es más que el proceso de asegurarse de que en cualquier ecuación, las unidades (también llamadas «dimensiones», aunque no tienen nada que ver con las direcciones en el espacio-tiempo) de dos cantidades cualesquiera que sumamos o igualamos tienen que coincidir. No puedes sumar un número de centímetros a un número de gramos.

La mayoría de las unidades físicas pueden considerarse combinaciones, mediante multiplicación o división, de un conjunto básico de tres: distancia [D], tiempo [T] y masa [M]. Consideremos algo como la energía. La fórmula newtoniana de la energía cinética es $E = \frac{1}{2} mv^2$. Se dice que el número puro ½ es «adimensional». Así pues, utilizando corchetes para denotar las «dimensiones de», tenemos

$$[\text{energía}] = [\text{masa}][\text{velocidad}]^2 = [\text{M}][\text{D}]^2/[\text{T}]^2, \qquad (6.4)$$

ya que la velocidad tiene unidades de distancia dividida por tiempo.

Hemos estado utilizando lo que los físicos de partículas llaman **unidades naturales**, en las que

$$\hbar = c = 1. \qquad (6.5)$$

Esto implica inmediatamente relaciones entre las distintas unidades básicas. Para la velocidad de la luz, tenemos

$$[c] = [\text{D}]/[\text{T}] = 1. \qquad (6.6)$$

Por consiguiente, cuando $c = 1$ podemos decir que la distancia y el tiempo se miden en las mismas unidades, [D] = [T]. Si lo desea, puede medir el tiempo en centímetros. Para volver a las unidades normales, basta con multiplicar o dividir por factores de c hasta obtener las unidades deseadas. Del mismo modo, para la constante reducida de Planck $\hbar = 1{,}05 \times 10^{-27}$ g cm^2 s^{-1} tenemos

$$[\hbar] = [M][D]^2/[T] = 1. \tag{6.7}$$

Esto es un poco más difícil de poner en pie rápidamente, pero la filosofía es la misma.

Cuando combinamos $c = 1$ con $\hbar = 1$, encontramos que las tres unidades básicas de masa/distancia/tiempo pueden reducirse a una única escala común. Los físicos de partículas suelen elegir la energía como escala, aunque utilicen «masa» y «energía» indistintamente. Ya hemos hecho uso de esto más arriba cuando expresamos las masas de las partículas en términos de electrón-voltios, que en realidad son unidades de energía. Abreviaturas comunes incluyen GeV para giga (10^9) eV, MeV para mega (10^6) eV y meV para mili (10^{-3}) eV. También se encontrará con keV para kilo (10^3) eV, TeV para tera (10^{12}) eV y μeV para micro (10^{-6}) eV.

Una aplicación adecuada de (6.6) y (6.7) revela que podemos expresar todas las unidades básicas en términos de energía como

$$[M] = [E], \quad [D] = 1/[E], \quad [T] = 1/[E]. \tag{6.8}$$

Así pues, las unidades de cada cantidad pueden considerarse energía elevada a algún exponente. La distancia y el tiempo no son más que energías inversas, coherentes con la relación física (6.3). (Cuando $c = 1$, el momento y la energía tienen las mismas unidades).

ANÁLISIS DIMENSIONAL

Ahora ya podemos realizar algunos de los análisis dimensionales prometidos anteriormente, empezando por pensar en los lagrangianos y averiguando las dimensiones de todo en unidades de energía elevada a una potencia apropiada. Nuestra definición original del lagrangiano L era la energía cinética menos la energía potencial, así que está claro que el propio lagrangiano tiene dimensiones de [E]. Pero en la teoría de campos utilizamos más a menudo la densidad de Lagrange, relacionada con el lagrangiano mediante

$$L = \int \mathcal{L}\, d^3x. \tag{6.9}$$

(Ahora denotamos explícitamente las tres dimensiones del espacio, ya que el número de dimensiones espaciales afectará a las unidades de las cosas). La notación d^3x se refiere a un elemento de volumen infinitesimal del espacio: el producto de longitud por anchura por altura. Por tanto, tiene unidades

$$[d^3x] = [\mathrm{D}]^3 = 1/[\mathrm{E}]^3. \tag{6.10}$$

El signo integral no tiene dimensiones. Así que, para obtener el lagrangiano, con unidades de [E], multiplicamos la densidad de Lagrange por el elemento volumen, que tiene unidades $1/[\mathrm{E}]^3$. Por lo tanto, la propia densidad de Lagrange debe tener unidades

$$[\mathcal{L}] = [\mathrm{E}]^4. \tag{6.11}$$

Consideremos una teoría de campo escalar simple, con densidad de Lagrange (5.4), es decir, densidades de energía cinética menos energía de gradiente menos energía potencial. Deseamos averiguar las unidades del campo ϕ. Podemos llegar a ello considerando cualquiera de los términos (se suman, por lo que deben tener las mismas unidades), así que veamos la densidad de energía cinética $1/2(\partial\phi/\partial t)^2$. Una derivada parcial con respecto al tiempo tiene unidades $1/[\mathrm{T}] = [\mathrm{E}]$

(porque el tiempo está en el denominador de $\partial/\partial t$), y queremos que todo el término tenga unidades de una densidad de Lagrange, $[E]^4$. Por tanto, el campo tiene dimensiones de energía,

$$[\phi] = [E]. \tag{6.12}$$

Esto es válido esencialmente para todos los campos escalares, aunque otros tipos de campos podrían tener dimensiones diferentes. Los electrones y otros campos fermiónicos, por ejemplo, tienen unidades $[E]^{3/2}$.

Estos son los dos hechos básicos que necesitamos: el campo escalar tiene unidades de [E], y la densidad de Lagrange tiene unidades de $[E]^4$. Cuando escribimos términos de interacción, cada uno viene con una cierta constante de acoplamiento, que nos indica la importancia del vértice correspondiente en un diagrama de Feynman. Cada constante de acoplamiento deberá tener unidades que hagan que el término de interacción en su conjunto tenga dimensiones de $[E]^4$, porque es un término en la densidad de Lagrange.

RENORMALIZABILIDAD

¿Por qué nos importan tanto las unidades de los campos y del lagrangiano y sus constantes de acoplamiento? Porque en una teoría de campo efectivo, esas «constantes» de acoplamiento no son realmente constantes, dependen del valor del corte UV Λ. Y esa dependencia será justamente lo que necesitamos para asegurarnos de que nuestra teoría efectiva nos dé las mismas predicciones, independientemente del corte que decidamos imponer. Mejor aún, implicará que muchas interacciones que podríamos escribir son, en última instancia, irrelevantes para hacer predicciones precisas.

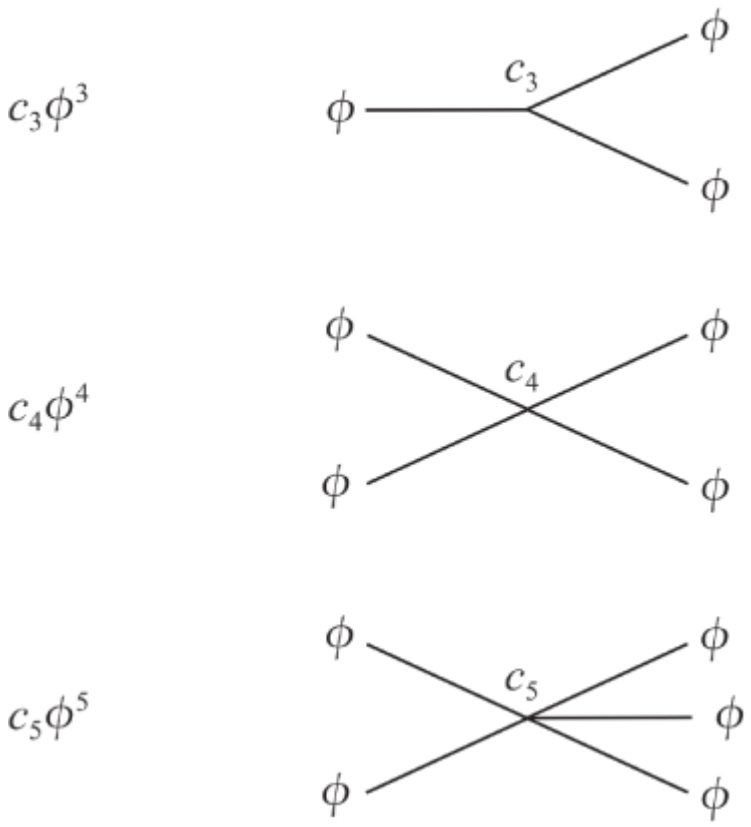
Para simplificar, consideremos una teoría de un campo escalar ϕ con una densidad de Lagrange de interacción formada por tres

términos diferentes:

$$\mathcal{L}_{int} = c_3\phi^3 + c_4\phi^4 + c_5\phi^5. \qquad (6.13)$$

Recuerde que esto es a lo que dedican mucho tiempo trabajando los teóricos del campo cuántico: empezar con una teoría de campo libre y luego escribir lagrangianos de interacción y pensar en sus consecuencias. Para entrar en la mentalidad de la física de partículas, vamos a observar detenidamente (6.13) durante un rato y pensar en lo que cada término representa físicamente, y en cómo se ve afectado cada uno de ellos por la renormalización.

Cada término tiene una interpretación física como un cierto número de partículas que interaccionan entre sí, como se manifiesta en los vértices que podemos unir para hacer diagramas de Feynman completos. Los términos simples de (6.13) dan lugar a vértices con tres, cuatro o cinco partículas implicadas, una partícula por cada aparición del propio campo. A cada término se le asocia su propia constante de acoplamiento.



Nuestras partículas escalares sin florituras no se distinguen de sus propias antipartículas, así que no importa si dibujamos las líneas como entrantes o salientes: hay un vértice correspondiente para cada disposición que podamos imaginar.

Tenemos que pensar en las unidades asociadas a estas constantes de acoplamiento, que, a su vez, nos dirán cómo se ven afectadas por nuestra elección del corte UV Λ. Dado que ϕ tiene unidades [E] y $\mathcal{L}_{int}$ de $[E]^4$, conocemos inmediatamente las unidades de nuestros acoplamientos

$$[c_3] = [E]\ , \quad [c_4] = 1\ , \quad [c_5] = 1/[E]. \tag{6.14}$$

Más generalmente, para un término de la forma $c_n\phi^n$, el acoplamiento tendrá unidades $[c_n] = [E]^{4-n}$. Ese 4 en el exponente es solamente la dimensionalidad del espacio-tiempo; en teorías

hipotéticas en dimensiones superiores, las constantes de acoplamiento tendrán unidades diferentes.

A mediados del siglo XX, cuando los físicos luchaban contra los infinitos en la TCC y antes de que se desarrollara el enfoque de la TCE, se dieron cuenta rápidamente de que el procedimiento de renormalización que funcionaba para la electrodinámica cuántica no funcionaría para todas las teorías. Algunas teorías son **renormalizables**, pero otras no. En una teoría renormalizable, los infinitos que obtenemos de los diagramas de bucle pueden cancelarse en términos de un pequeño número de parámetros, lo que nos deja con una bonita teoría predictiva. La EDC es un ejemplo paradigmático.

Y en su mayor parte, la distinción es sencilla: en una teoría renormalizable, todas las constantes de acoplamiento tienen unidades $[\mathrm{E}]^a$ con $a \geq 0$. Sus dimensiones son de energía elevada a alguna potencia no negativa. En nuestro simple modelo escalar, c_3 y c_4 estarían permitidos, pero tendríamos que establecer $c_5 = 0$ para que nuestra teoría fuera renormalizable.

A grandes rasgos, entonces, los términos lagrangianos con un número pequeño de campos son renormalizables, mientras que los que tienen un número grande de campos no lo son. Y no hay tantos términos que se puedan escribir con un número pequeño de campos. Así que la exigencia de que una teoría fuera renormalizable era realmente útil para el trabajo del físico: solo una clase muy restringida de posibles teorías cumpliría los requisitos.

Una vez que se incluyen términos de interacción con acoplamientos cuyas dimensiones son potencias negativas de la energía —lo que ocurre en nuestro actual ejemplo, cuando hay más de cuatro apariciones del campo escalar, como el término $c_5\phi^5$— la teoría es **no renormalizable**. Cuando las constantes de acoplamiento tienen unidades de potencias negativas de energía, la expresión resultante para los diagramas de Feynman necesita tener potencias positivas del momento, para que las dimensiones globales sean correctas. Así que

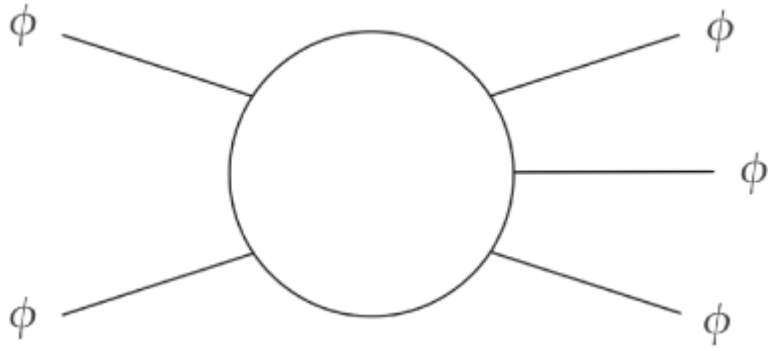
cuando integramos hasta un momento grande, es fácil llegar a infinito. De hecho, por lo general habrá un número infinito de infinitos, cada uno de los cuales requiere un tratamiento especial: la introducción de un nuevo término para cancelar cada infinito de la manera precisamente correcta. No podemos medir físicamente un número infinito de observables para determinar las características de todos estos nuevos términos. Así pues, el problema de las teorías no renormalizables no es que estén fundamentalmente definidas de forma pobre o incorrecta, sino que requieren un número infinito de datos para poder hacer predicciones experimentales, lo que parece una mala cualidad para una teoría física.

En su momento, los físicos respondieron a esta situación insistiendo en que solo debían considerar teorías renormalizables. Esto fue de gran ayuda, ya que redujo el número potencialmente infinito de términos que se podían introducir en un lagrangiano a un número manejable. En la década de 1960, cuando se propuso por primera vez la idea de un bosón de Higgs y se conjeturó que podría ayudar a unificar las fuerzas electromagnética y nuclear débil, la mayoría de los físicos no prestaron mucha atención porque pensaban que la teoría no sería renormalizable. Más tarde, a principios de la década de 1970, Gerard 't Hooft y Martinus Veltman demostraron que la teoría electrodébil era renormalizable al fin y al cabo. De repente, todo el mundo se dejó llevar por el entusiasmo y empezaron a tomarse la teoría en serio. Todas las nuevas partículas que predecía acabaron siendo descubiertas, y 't Hooft y Veltman ganaron en 1999 el Premio Nobel de Física por su trabajo.

LAGRANGIANOS EFECTIVOS

El paradigma de la teoría de campo efectivo distorsiona un poco esta perspectiva. Imaginemos que tenemos nuestra teoría de campo escalar con interacciones (6.13), pero definimos $c_5 = 0$ para que la teoría sea

renormalizable. Incluso sin ese vértice fundamental de cinco partículas, seguirá habiendo procesos en los que intervengan cinco partículas reales, con (por ejemplo) dos partículas entrantes y tres salientes. Podríamos combinar vértices de tres partículas para formar un diagrama de un bucle con cinco partículas externas en total, como se muestra en la figura.



Desde la perspectiva de la TCE, esto da lugar a una interacción *efectiva* de cinco partículas. Las reglas del juego son que construimos un **lagrangiano efectivo** completo, donde partimos del lagrangiano fundamental (renormalizable) y luego, en lugar de hacer lo de siempre, integrar los momentos de bucle hasta el final y restar los infinitos, ajustamos los valores de las constantes de acoplamiento para que incluyan todos los efectos que habrían inducido. Por lo tanto, no es exacto decir que los momentos de bucle arbitrariamente grandes simplemente se descartan; decimos que quedan **integrados** y que sus efectos aparecen en las constantes de acoplamiento efectivas. Estamos agrupando procesos UV desconocidos en un simple conjunto de parámetros IR.

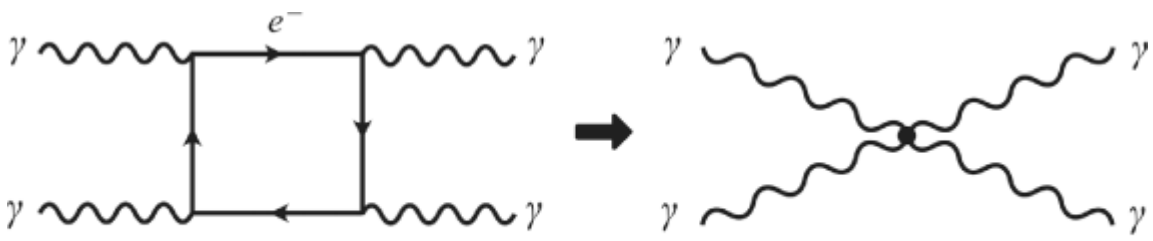
El lagrangiano efectivo no solamente incluye los términos renormalizables con los que empezamos; incluye todos los términos que se nos ocurran. Por tanto, tendrá un término de la forma $c_5\phi^5$, y de hecho todos los términos posibles $c_n\phi^n$ con n más altos, para desgracia de la renormalizabilidad. Pero la aparente no renormalizabilidad no nos molesta, porque no usamos ese lagrangiano para integrar los

momentos de bucle hasta el infinito; integramos solamente hasta el corte UV Λ, y siempre se obtienen respuestas finitas.

Este procedimiento ayuda a responder a una pregunta que puede haber estado atormentándole durante un tiempo. Inventamos una teoría de campo efectivo introduciendo un corte ultravioleta Λ, y luego no incluyendo los momentos UV por encima de esa escala cuando hicimos integrales de bucle en nuestros diagramas de Feynman. Pero ¿cómo elegimos cuál debe ser Λ? Si esa elección afecta a las contribuciones que tenemos en nuestros diagramas, parece que va a tener un gran impacto sobre lo que predice nuestra teoría.

Esa suposición es errónea. La elección de Λ no tiene ningún efecto sobre nuestras predicciones teóricas. Las partículas virtuales UV pueden tener un efecto, pero todos esos efectos están agrupados en los valores de las constantes de acoplamiento en el lagrangiano efectivo. Así es como llegamos a una teoría que solamente necesita hablar de partículas IR.

Para poner un poco de carne en estos huesos: la masa del fotón es 0 y la masa del electrón es 0,511 MeV (millones de electrón-voltios). Supongamos que fijamos nuestro corte ultravioleta en 1 keV (mil electrón-voltios), muy por debajo de la masa del electrón. Entonces el electrón es una partícula UV, y tenemos una teoría en la que solamente hay fotones. El lagrangiano para la EDC no incluye ningún acoplamiento directo entre fotones y otros fotones, solamente entre fotones y pares electrón/positrón. Pero podemos escribir una TCE con un corte de 1 keV que incluya fotones que interaccionen entre sí; esas interacciones no renormalizables son inducidas por bucles de electrones a energías más altas. (Heisenberg y su colega Hans Heinrich Euler calcularon en realidad este efecto en 1936, sin las ventajas de los diagramas de Feynman o de la teoría de campo efectivo). Esta interacción efectiva de cuatro fotones aparece como un vértice básico en la TCE, que resume un montón de procesos UV que no estamos calculando explícitamente.



Electrodinámica cuántica TCE con $\Lambda = 1$ keV

EL GRUPO DE RENORMALIZACIÓN

Para que todo esto funcione, tenemos que reconocer que las «constantes» de acoplamiento que aparecen en el lagrangiano de la TCE no son realmente constantes en absoluto. Son funciones de Λ. Eso tiene sentido. El significado de una constante de acoplamiento en la densidad de Lagrange es simplemente la «fuerza» de la interacción correspondiente. La constante de estructura fina α, por ejemplo, nos indica la fuerza de la interacción electromagnética. Y en una teoría de campo efectivo con corte Λ, estamos agrupando todos los efectos de partículas virtuales con energías *superiores* a Λ en el valor del acoplamiento efectivo. Así que, por supuesto, el acoplamiento efectivo dependerá del valor de Λ, ya que gobierna el comportamiento de los procesos UV que estamos agrupando.

Por lo tanto, decimos que las constantes de acoplamiento «fluyen» o «se mueven» en función del corte, y nos referimos a ellas como **constantes de acoplamiento variables**. En lugar de escribir simplemente c_4 como el coeficiente de ϕ^4 en nuestra teoría de campo escalar, lo escribiríamos como una función del corte, $c_4(\Lambda)$. La forma en que las constantes de acoplamiento cambian para diferentes valores del punto de corte sigue una técnica que se denomina **grupo de renormalización**.

De hecho, podemos hacer una conjetura decente sobre cómo evolucionarán nuestras constantes de acoplamiento a medida que cambiamos Λ, simplemente utilizando el análisis dimensional. Ya

vimos en (6.14) cómo los diferentes acoplamientos tienen diferentes unidades, y sabemos que los acoplamientos efectivos son funciones de Λ. Así que la idea más simple, que resulta ser bastante exacta, es que los acoplamientos son proporcionales a potencias adecuadas de Λ:

$$c_3(\Lambda) \propto \Lambda, \quad c_4(\Lambda) \propto \Lambda^0, \quad c_5(\Lambda) \propto \frac{1}{\Lambda}, \ldots \tag{6.15}$$

Vemos un patrón obvio. Si nuestra interacción ϕ^n tiene $n < 4$, entonces su constante de acoplamiento será proporcional a una potencia positiva del corte Λ. A medida que aumentemos los valores de Λ, el acoplamiento será cada vez mayor. Nos referimos a estos términos de interacción como **operadores relevantes**, ya que los campos y sus combinaciones pueden ser considerados operadores en el espacio de Hilbert de la teoría cuántica de campos. Cuando tenemos interacciones en las que intervienen más campos, ϕ^n con $n > 4$, el acoplamiento evolucionará con una potencia negativa de Λ, y entonces hablamos de **operadores irrelevantes**. Se hacen más pequeños a medida que Λ crece.

Para ϕ^n con $n = 4$, el acoplamiento es adimensional, y tales términos se denominan **operadores marginales**. Un operador marginal puede evolucionar con Λ, pero lo hará de forma más gradual —normalmente como logaritmo de Λ— que la simple ley exponencial que obtenemos para los operadores relevantes e irrelevantes. No es obvio a primera vista si un operador marginal aumentará o disminuirá a medida que aumentamos el corte; hay que hacer un cálculo explícito para averiguarlo. En electrodinámica cuántica, la constante de estructura fina α constituye un acoplamiento adimensional que crece a altas energías. De hecho, se hace infinitamente grande a una escala superalta conocida como el **polo de Landau**, llamado así por el físico soviético Lev Landau. Esto es un problema, y es una de las razones por las que mucha gente piensa que la EDC necesita ser incluida en última instancia en una teoría unificada que se comporte mejor. En

cambio, en la cromodinámica cuántica (CDC), que es la teoría de la fuerza nuclear fuerte, el acoplamiento disminuye a altas energías. Esto se conoce como **libertad asintótica**, ya que los quarks y los gluones pasan a ser libres (no interaccionan) a medida que aumenta la energía. La CDC parece ser una teoría cuántica de campos con un comportamiento perfecto hasta energías arbitrariamente altas.

Todas las designaciones de relevante/marginal/irrelevante para los operadores (términos en el lagrangiano) tienen sentido. La elección del corte UV depende de nosotros, con una regla: tiene que estar a mayor energía que cualquiera de las masas o momentos de las partículas que estemos considerando en ese instante. Es decir, Λ debe ser grande en comparación con las escalas de interés de nuestro problema. A medida que consideramos valores cada vez mayores, los acoplamientos correspondientes a los operadores relevantes crecerán, los de los operadores irrelevantes se reducirán y los de los operadores marginales no cambiarán mucho. Por tanto, los operadores irrelevantes a menudo pueden despreciarse, al menos en una primera aproximación.

Esto explica por qué las teorías renormalizables parecían dignas de especial atención en primer lugar. No sabemos cuál es la teoría última de todo, ni siquiera si es una teoría cuántica de campos, desde luego no a escalas ultrapequeñas y energías altas. Pero una teoría de campo efectivo que se derive de esa teoría última *parecerá* generalmente renormalizable, al menos hasta una buena aproximación. Los efectos de los momentos de bucle UV tienden a hacer que los términos renormalizables (relevantes y marginales) parezcan grandes, y los términos no renormalizables (irrelevantes) parezcan pequeños. Puede ocurrir todo tipo de locuras a distancias mínimas, y el mundo seguirá pareciendo correcto y renormalizable ante nosotros.

La siguiente lista resume la situación. (Únicamente incluimos términos con un número par de campos, para simplificar las cosas). En la versión anticuada de la teoría cuántica de campos, nos restringiríamos solamente a los términos renormalizables,

integraríamos los momentos de bucle hasta escalas arbitrariamente altas y sustraeríamos los infinitos resultantes mediante renormalización. En un enfoque moderno de teoría de campo efectivo, permitimos todos los términos que podemos escribir, pero solamente integramos los momentos de bucle hasta un corte UV Λ, que mantiene todo finito, y los acoplamientos dependen de Λ para mantener las predicciones físicas independientes de Λ. Los términos con muchos campos generalmente tienen acoplamientos irrelevantes que se hacen diminutos a medida que Λ se hace grande, así que no tenemos que preocuparnos mucho por ellos de todos modos.

— Teoría renormalizable de campos escalares:

$$\mathcal{L} = \frac{1}{2}(\partial_t \phi)^2 - \frac{1}{2}(\partial_x \phi)^2 - \frac{1}{2}m^2\phi^2 + c_4\phi^4$$

  — las partículas virtuales tienen energías arbitrariamente altas
  — se restan los infinitos mediante renormalización

— Teoría del campo escalar efectivo:

$$\mathcal{L} = \frac{1}{2}(\partial_t \phi)^2 - \frac{1}{2}(\partial_x \phi)^2 - \frac{1}{2}m^2(\lambda)\phi^2 + c_4(\lambda)\phi^4 + c_6(\lambda)\phi^6 + \cdots$$

  — las partículas virtuales tienen energías por debajo del corte UV Λ
  — las constantes de acoplamiento dependen de Λ para que las predicciones físicas no lo hagan

Un detalle molesto que merece la pena mencionar: puede parecer desconcertante que algo tan directo y mensurable como la masa de la partícula pueda depender del corte. No le falta motivo para el desconcierto. La resolución es que en esta visión perdemos la bonita equivalencia entre la masa de la partícula y el parámetro m que aparece en el lagrangiano. Sigue habiendo una masa «física» de la

partícula; no es igual a m, pero podemos calcularla como una combinación de $m(\Lambda)$ y los acoplamientos de orden superior como $c_4(\Lambda)$ y así sucesivamente. Los físicos se han reconciliado con el hecho de que algunas cosas son simplemente complicadas.

MISTERIOS DE LA NATURALIDAD

El paradigma de la teoría de campo efectivo es convincente intelectualmente y útil en la práctica. Desempeña un papel central en la física de partículas moderna y se ha extendido a contextos tan diversos como el crecimiento de estructuras en cosmología y la generación de ondas gravitacionales por parte de los inspiradores agujeros negros.

Pero hay preocupaciones que acechan. Estas se centran en la idea de la **naturalidad** de los parámetros en física de partículas. La naturalidad no es un principio rigurosamente formulado, pero motiva muchas investigaciones en la teoría de partículas contemporánea. La idea básica es que los parámetros numéricos en un lagrangiano efectivo no deberían ser muy pequeños comparados con la unidad, a menos que haya una simetría u otra buena razón para que aparezca un número tan reducido.

Lo que tenemos en mente es que, pensando en los diagramas de Feynman, los parámetros efectivos de baja energía que medimos son en realidad la suma de muchas contribuciones diferentes. Algunas contribuciones podrían ser tanto negativas como positivas. Sería natural esperar que la suma de todas estas contribuciones fuera del mismo orden que las contribuciones individuales. En otras palabras, si sumáramos un montón de números de diez cifras (algunos positivos, otros negativos), sería extraño que el resultado fuera exactamente cero, o incluso cualquier número entre –10 y +10. Esto podría suceder conceptualmente, pero en apariencia sería una coincidencia inusual de alguna clase, a menudo referida como **ajuste fino**. (Nada que ver con

el ajuste fino de parámetros para permitir la existencia de vida, que también se oye debatir en discusiones cosmológicas/teológicas).

Y, sin embargo, esto parece ocurrir de hecho en el mundo real, en dos casos bien conocidos. Uno es la masa del bosón de Higgs, m_H = 125 GeV (giga electrón-voltios = mil millones de eV). Puede parecer un gran número, más de cien veces la masa del protón. Pero si etiquetamos el campo de Higgs con H, el término de masa en el lagrangiano efectivo es simplemente ½ $m_H^2 H^2$. Es un operador relevante, y por análisis dimensional esperaríamos que el acoplamiento m_H^2 evolucionara como Λ^2. Hay que reconocer que es un poco extraño, ya que se supone que el corte UV Λ representa un parámetro que introdujimos para nuestra propia conveniencia, no una verdadera constante de la naturaleza. Pero, en este contexto, simplemente nos recuerda que los diagramas de bucle tienen el efecto de aumentar el valor eficaz de la masa de Higgs, en la medida de lo posible. La nueva física podría empezar a hacer efecto a una escala de alta energía, como la escala de Planck, $M_P = 10^{18}$ GeV. Pero entonces esperaríamos que la masa de Higgs medida estuviera cerca de la escala de Planck, y no lo está. Difieren bastante:

$$\frac{m_H}{M_P} \sim 10^{-16}. \tag{6.16}$$

Este es exactamente el tipo de número pequeño que el principio de naturalidad parecería prohibir, pero ahí está. Esta discrepancia entre la expectativa y la realidad se conoce como el **problema de la jerarquía** en la física de partículas. Podría resolverse si, por ejemplo, hubiera nuevos efectos físicos cerca de una escala de energía de 100 GeV que funcionaran para estabilizar la masa de Higgs. Un ejemplo es la supersimetría —una simetría hipotética entre bosones y fermiones— que podría actuar anulando las diferentes contribuciones de los bucles entre sí. Ha habido otras propuestas, pero en general todas predicen que deberíamos ver nuevas partículas cerca de esta escala de energía.

La búsqueda de tales partículas fue una de las principales motivaciones para construir el Gran Colisionador de Hadrones, pero hasta el momento de escribir estas líneas no se han encontrado tales partículas. El problema de la jerarquía sigue sin resolverse. (Otras partículas tienen masas, por supuesto, pero en todos los casos conocidos aparte del Higgs existe una simetría u otro mecanismo que explique por qué sus masas son tan pequeñas en comparación con la escala de Planck).

Un escollo similar lo constituye el **problema de la constante cosmológica**. Como se señala brevemente en *Espacio, tiempo y movimiento*, la constante cosmológica fue introducida por Einstein como una modificación de su ecuación de la relatividad general. Desde la perspectiva de la teoría de campos, se puede considerar como el término más simple posible que podríamos incluir en un lagrangiano efectivo: una constante independiente ρ_0, que no multiplica a ningún campo en absoluto. Podemos interpretar que ρ_0 representa la densidad de energía del espacio vacío: la **energía del vacío**. La energía del vacío y la constante cosmológica son dos etiquetas diferentes para decir exactamente lo mismo: no deje que nadie le diga lo contrario.

Por análisis dimensional, tenemos que $[\rho_0] = [E]^4$. Desde el punto de vista de la TCE con corte Λ, esperaríamos que fuera del orden $\rho_0 \sim \Lambda^4$. Si no tenemos procesos físicos novedosos que pudieran empezar a hacer efecto anulando las contribuciones de los diagramas de bucle, podríamos esperar introducir aquí la masa de Planck M_P una vez más como valor razonable para el corte Λ. Esta vez nuestras expectativas son aún más dramáticamente erróneas que en el caso del problema de jerarquía. Por observación, tenemos

$$\frac{\rho_0}{M_P^4} \sim 10^{-122}. \tag{6.17}$$

Se trata de una gran discrepancia, famosa por ser el mayor desajuste entre expectativa y observación de toda la física teórica.

Tanto el problema de la jerarquía como el de la constante cosmológica son, con todas las letras, fallos bastante espectaculares de nuestra intuición de la TCE. Hasta la fecha no hay consenso sobre cómo abordarlos. No faltan propuestas, pero ninguna de ellas ha calado realmente. No tenemos por qué desechar todo el contenido de la TCE, pero tal vez existan formas sutiles por las que los efectos UV están llegando a ajustar nuestros parámetros IR de formas imprevistas. O quizás esté ocurriendo algo incluso más dramático. Es precisamente cuando nuestras teorías no concuerdan del todo con nuestros datos que la ciencia se vuelve apasionante.

---

1 Se cita a Dirac en H. Kragh, *Dirac: A Scientific Biography* (Cambridge University Press, 1990), 184. Las citas de Feynman proceden de R. P. Feynman, *QED: The Strange Theory of Light and Matter* (Penguin, 1990), 128 (N. de la T.: Editado en castellano por Alianza Editorial en 1988 bajo el título *Electrodinámica cuántica: La extraña teoría de la luz y la materia*).

SIETE

ESCALA

La evocadora película *Powers of Ten* de Charles y Ray Eames (basada en un libro anterior de Kees Boeke) invita al espectador a experimentar el alcance del universo partiendo de una vista de un solo metro cuadrado —una pareja disfrutando de una comida campestre en la orilla del lago Chicago— y ampliando la imagen en un factor de diez cada diez segundos. Vemos una ciudad que se encoge de manera gradual, y después la Tierra, el sistema solar, estrellas cercanas, la galaxia de la Vía Láctea, y finalmente cúmulos de galaxias y estructuras aún mayores. A continuación, volvemos a acercarnos e iniciamos otro viaje a escalas más profundas, viendo células de la piel, orgánulos, moléculas, átomos y partículas elementales.

Hay sustancias, como la materia oscura o los neutrinos de fondo, que no vemos directamente en este tipo de imágenes. Todo lo que vemos es un átomo o bien está hecho de átomos. Y los átomos están formados por solamente tres constituyentes: protones, neutrones y electrones. A veces subestimamos la enorme jerarquía de escala que constituye el viaje desde las partículas elementales hasta el universo.

Nuestros últimos capítulos han sido una desafiante visión general de ideas teóricas profundas. En este capítulo vamos a tomarnos un respiro considerando un tema que es más fácil de visualizar: de qué está hecho realmente el universo conocido y, en particular, cuáles son las distintas escalas de masa y energía que caracterizan nuestro mundo.

REPASO DE LAS UNIDADES

Hemos estado utilizando «unidades naturales», en las que $\hbar = c = 1$. Además, tenemos las siguientes relaciones entre las dimensiones de energía, masa, distancia y el tiempo:

$$[\mathrm{E}] = [\mathrm{M}] = 1/[\mathrm{D}] = 1/[\mathrm{T}]. \tag{7.1}$$

Una vez hecha esta elección, todavía tenemos que elegir una unidad fundamental para usar, y los físicos de partículas suelen elegir la energía. En particular, medirán la energía en **electrón-voltios** (eV), definida como la energía cinética que ganaría un electrón al acelerar debido a una diferencia de potencial eléctrico de un voltio. La mayoría de las personas, incluidos los físicos de partículas profesionales, no tienen mucha intuición física para lo que es empujar electrones individuales a través de diferencias de potencial. Pero otras unidades de energía no son mucho más intuitivas. El consumo doméstico de energía suele medirse en kilovatios-hora (kWh) o unidades térmicas británicas (BTU). La energía almacenada en los alimentos se mide habitualmente en calorías, aunque las cifras que se encuentran en las directrices dietéticas o en los listados nutricionales son en realidad kilocalorías (kcal). En el sistema métrico podemos utilizar julios (equivalentes a kg m^2/s^2) o ergios (g cm^2/s^2). Las siguientes relaciones permiten realizar conversiones de uno a otro sistema.

$$\begin{aligned} 1\ \mathrm{eV} &= 1{,}60\times10^{-19}\ \mathrm{J} = 1{,}60\times10^{-12}\ \mathrm{ergios} = 4{,}45\times10^{-26}\ \mathrm{kWh} = \\ &= 1{,}52\times10^{-22}\ \mathrm{BTU} = 3{,}83\times10^{-23}\ \mathrm{kcal}. \end{aligned} \tag{7.2}$$

Vemos al menos una característica conveniente del electrón-voltio: es una cantidad diminuta de energía. Las partículas son cosas diminutas, así que este tipo de unidades tienen sentido.

Pensando como un físico de partículas, podemos convertir unidades comunes a electrón-voltios y viceversa.

$$1\ \text{eV} \approx 10^{-33}\ \text{gramos},$$
$$1\ \text{eV}^{-1} \approx 10^{-5}\ \text{cm},$$
$$1\ \text{eV}^{-1} \approx 10^{-15}\ \text{s}. \qquad (7.3)$$

No debemos interpretar estas relaciones como si pudiéramos transformar físicamente una cantidad de distancia en una cantidad de energía, ni nada por el estilo. Se trata simplemente de distintas elecciones de unidades para medir las cosas. Pero, como veremos, proporcionan reglas generales para relacionar distancias y tiempos comunes con procesos de física de partículas. Un electrón-voltio es una pequeña cantidad de masa, y su inverso es una pequeña distancia y un corto periodo de tiempo, al menos para los estándares humanos.

LAS ESCALAS DE LA FÍSICA DE PARTÍCULAS

Casi ningún físico de partículas podría decirle la masa del electrón en gramos. Pero casi seguro que sería capaz de indicar su masa en eV, junto con la de otras partículas conocidas.

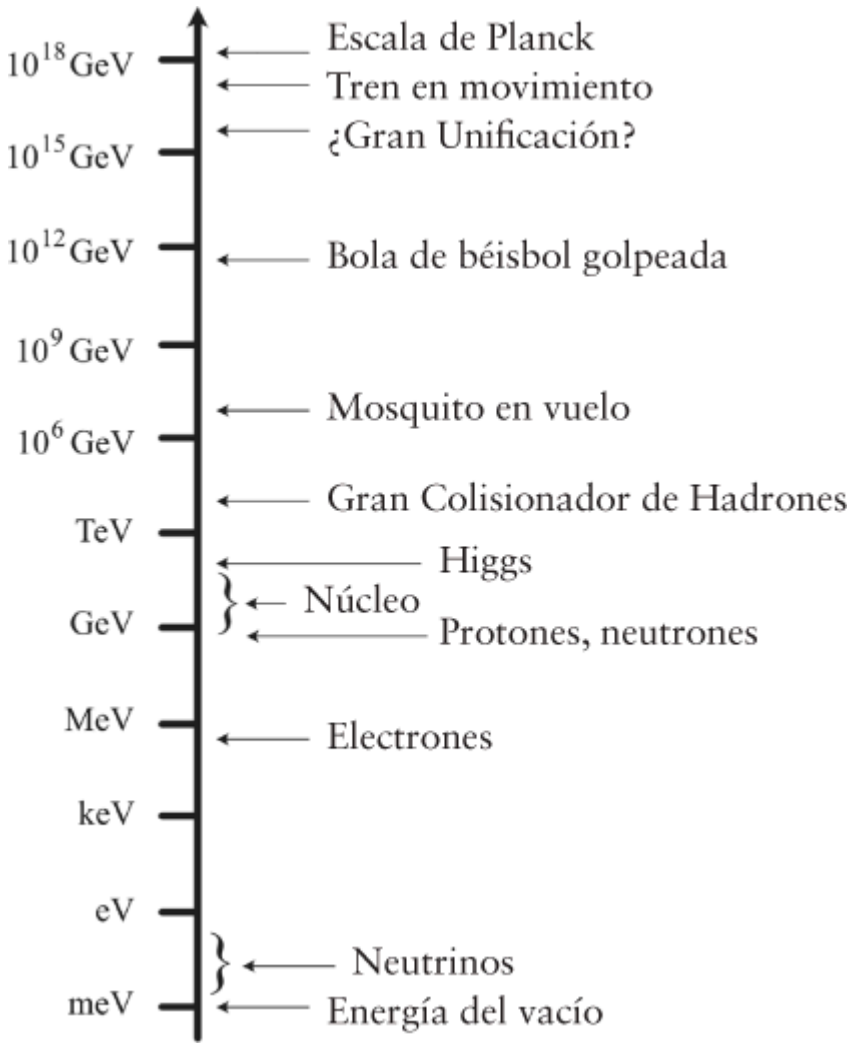
Electrón	$5{,}11\times10^{5}$ eV	0,511 MeV
Protón	$9{,}383\times10^{8}$ eV	0,9383 GeV
Neutrón	$9{,}396\times10^{8}$ eV	0,9396 GeV
Neutrino	$\leq 10^{-1}$ eV	≤ 100 meV

Un GeV es una unidad cómoda de usar para la física de partículas, ya que es aproximadamente la masa de un protón o un neutrón. Hemos incluido aquí el neutrino, que ya hemos mencionado, pero del que no hemos hablado realmente en detalle. Los neutrinos son partículas de poca masa, neutras y casi invisibles que resultan tener propiedades extraordinariamente sutiles.

Para orientarnos, resulta útil representar estas y otras escalas de la física de partículas en una escala logarítmica. Hemos incluido no

solamente las masas de las partículas anteriores, sino también la masa reducida de Planck ($\sqrt{\hbar c / 8\pi G}$, donde G es la constante gravitatoria de Newton), las energías de colisión típicas del Gran Colisionador de Hadrones, la masa del bosón de Higgs, los núcleos atómicos y la escala de la energía del vacío, E_{vac}. Para definir esta última, observamos que la densidad de energía del vacío ρ_0 tiene unidades $[E]^4$, y definimos $E_{vac} = \rho_0 ¼$. Por último, incluimos la hipotética escala de energía de la **Gran Unificación**, un escenario que unificaría las interacciones fuerte, débil y electromagnética (pero no la gravedad).

A modo de comparación, hemos incluido las energías cinéticas de algunos fenómenos cotidianos: un mosquito en vuelo, una pelota de béisbol bateada y un tren que se mueve a gran velocidad. (Para las partículas enumeradas citamos energías en reposo, es decir, masas, no energías cinéticas). Estas energías cotidianas son generalmente superiores a las energías del Gran Colisionador de Hadrones y otros aceleradores de alta energía, lo que puede resultar sorprendente. ¿Por qué tenemos que gastar tanto dinero en una máquina gigantesca que ni siquiera puede igualar la energía de un mosquito? La respuesta, por supuesto, es que la energía de colisión del LHC se concentra en solamente dos protones. Los mosquitos y otros objetos macroscópicos tienen más energía, pero esto es porque están formados por muchas partículas; su energía por partícula es relativamente baja.



Cuando los científicos dicen «muchas partículas», es útil que tengan en mente un orden de magnitud aproximado. Una elección conveniente es el **número de Avogadro**, $N_A = 6{,}02\times10^{23}$. Ese es aproximadamente el número de **nucleones** (protones y neutrones) en un gramo de materia atómica. (Cuando se trata de átomos, los electrones son casi irrelevantes para su masa). El comportamiento macroscópico clásico puede empezar a hacer efecto en cantidades mucho más pequeñas, pero el número de Avogadro ofrece un buen punto de partida para relacionar los átomos con los objetos a escala humana. Un gramo es bastante pequeño, aproximadamente la masa de un clip de oficina. He aquí otras escalas familiares para comparar:

— Bacteria o ADN humano ~ 10^{12} nucleones
— Persona ~ 10^{28} nucleones
— Tierra ~ 10^{51} nucleones
— Sol ~ 10^{57} nucleones
— Galaxia de la Vía Láctea ~ 10^{70} nucleones
— Universo observable ~ 10^{80} nucleones

Es interesante contemplar estos números en sí mismos. Nuestro universo es una maraña de jerarquías, con estructuras coherentes de tamaños muy diferentes. En este capítulo nos contentaremos con pensar en los tamaños de los átomos individuales.

EL TAMAÑO DE UNA PARTÍCULA CUÁNTICA

Con estos órdenes de magnitud en mente, podemos pensar en volver a vincularlos con el mundo de la mecánica cuántica y la teoría cuántica de campos. Para ello, relacionaremos la masa y la distancia a través de un concepto crucial: la **longitud de onda Compton** de una partícula. Representa el espacio mínimo que puede ocupar una partícula.

Cuando Louis de Broglie sugirió por primera vez en 1924 que las partículas tenían propiedades ondulatorias, propuso la longitud de onda relevante que había que tener en cuenta para una partícula con momento p, lo que ahora llamamos la longitud de onda de De Broglie,

$$\lambda_{dB} = 2\pi/p. \tag{7.4}$$

(Antes lo escribimos como h/p, pero ahora establecemos $\hbar = h/2\pi = 1$). Esta es la escala de longitud que asociamos al comportamiento ondulatorio de la función de onda cuántica de una partícula masiva. Cuando calculamos la distancia entre bandas de interferencia en el experimento de la doble rendija, o exigimos que una órbita atómica

contenga un número entero de longitudes de onda, es la longitud de onda de De Broglie lo que tenemos en mente.

La longitud de onda Compton desempeña un papel diferente. Se define para una partícula de masa m mediante

$$\lambda_C = 1/m. \tag{7.5}$$

(Esta es la longitud de onda Compton «reducida»; la «normal» es esta multiplicada por 2π). A diferencia de la longitud de onda de De Broglie, la longitud de onda Compton no tiene nada que ver con el movimiento de la partícula; es una constante fija una vez que conocemos la masa. La longitud de onda Compton para un electrón es de aproximadamente 2×10^{-10} cm, mientras que para un protón es de aproximadamente 1×10^{-13} cm. La noción fue introducida por Arthur Compton en 1923 en su investigación sobre la dispersión fotón-electrón. Pero donde adquiere verdadero protagonismo es en la teoría cuántica de campos, donde expresa la dificultad para determinar la extensión espacial de una partícula.

Según la relatividad, la energía está relacionada con el momento y la masa de una partícula mediante

$$E^2 = p^2 + m^2. \tag{7.6}$$

Así, una partícula en reposo, con $p = 0$, tiene $E = m$, que reconocemos como una famosa ecuación una vez que recordamos que estamos definiendo $c = 1$. Pero también tenemos el principio de incertidumbre de Heisenberg, según el cual la imprecisión en las expresiones de posición y momento para la función de onda de una partícula debe satisfacer

$$\Delta x \cdot \Delta p \geq 1. \tag{7.7}$$

Imaginemos que tenemos una partícula cuya función de onda está localizada espacialmente en una región del tamaño de su longitud de onda Compton, de forma que

$$\Delta x \sim 1/m. \tag{7.8}$$

(Recuerde que ~ significa «es del orden de», por lo que en relaciones como esta ignoramos los factores numéricos del orden de la unidad, es decir, entre 0,1 y 10). Entonces no podemos simplemente decir que el momento es 0, ya que estará afectado de alguna incertidumbre, del orden

$$\Delta p \sim m. \tag{7.9}$$

Debido a esta incertidumbre en el momento, a partir de la ecuación (7.6) también habrá alguna incertidumbre en la energía, incluso si el valor medio del momento es $p = 0$:

$$\Delta E \sim \Delta p \sim m. \tag{7.10}$$

Así que, si tratamos de localizar una partícula dentro de su longitud de onda Compton, la incertidumbre energética tiene una magnitud $\Delta E \sim m$. Pero eso es suficiente energía para crear una partícula adicional completa, un proceso al cual la TCC está encantada de dar cabida. Y fíjese en que, si localizáramos la partícula en una región aún más pequeña, la incertidumbre en la energía sería aún mayor.

En el marco de la teoría cuántica de campos, esto significa que, si intentamos comprimir una partícula para que ocupe una región más pequeña que $\lambda_C = 1/m$, su estado cuántico pasa a ser una superposición del estado de una o más partículas, y no de una única partícula simplemente. La longitud de onda Compton es la escala de longitud más pequeña en la que podemos pensar con seguridad que el sistema

representa una sola partícula. En la práctica, esto significa que la longitud de onda Compton puede considerarse como el «tamaño» de una partícula, en lo que respecta a la TCC. Y nótese que disminuye a medida que aumenta la masa. En la teoría cuántica de campos, las partículas más pesadas ocupan menos espacio. Se trata de un concepto curioso del tamaño que, sin embargo, no concuerda del todo con nuestra intuición clásica. La función de onda de una partícula puede extenderse con facilidad en una distancia mayor que su longitud de onda Compton, pero no puede comprimirse en menos.

Hay una implicación inmediata y triste de este hecho: el Hombre Hormiga, el superhéroe que puede encogerse hasta tamaños arbitrariamente pequeños, no es realmente posible. Ni tampoco podemos imaginar que existan organismos o civilizaciones enteras dentro de un solo átomo. Si intentáramos reducir el número de partículas de Avogadro al tamaño de un protón, o bien cada una de esas partículas tendría que ser más masiva que el propio protón —en cuyo caso nos daríamos cuenta, porque la cosa sería mucho más pesada— o bien serían de menor masa, y sus longitudes de onda Compton serían demasiado grandes para caber en un espacio tan pequeño. A pesar de todas sus sorpresas, el reino cuántico microscópico es relativamente continuo y monótono. Las oscilaciones pequeñas transportan demasiada energía.

QUARKS, HADRONES, BARIONES, MESONES

La idea de que no podemos comprimir un conjunto de muchas partículas hasta escalas subatómicas puede parecer que se contradice con otro hecho que puede haber oído: cada protón y neutrón está formado por tres **quarks**. En la década de 1960 Murray Gell-Mann y (de forma independiente) George Zweig se dieron cuenta de que se podía dar sentido al conjunto rápidamente creciente de partículas que interaccionan fuertemente que se estaba descubriendo en los

aceleradores de partículas mediante el planteamiento de que todas ellas estaban formadas por componentes más pequeños, denominados quarks por Gell-Mann a partir de una frase del *Ulises* de James Joyce: «¡Tres quarks para Muster Mark!». La inspiración para el nombre vino de la idea de que los protones y neutrones, y muchos de sus primos hadrónicos, contenían tres quarks cada uno. Es de suponer que los quarks tienen menos masa que los protones y neutrones que forman al combinarse y, por tanto, mayores longitudes de onda Compton. Entonces, ¿cómo se pueden comprimir en el interior? Este es un buen momento para hablar un poco de los quarks y de las partículas que forman, llamadas **hadrones**.

Sí, los protones y los neutrones están formados por quarks, y también por los bosones sin masa portadores de la fuerza nuclear fuerte que los mantiene unidos, llamados **gluones**. Se conocen seis tipos de **sabores** para los quarks: ***up*** (arriba), ***down*** (abajo), ***charm*** (encanto), ***strange*** (extraño), ***top*** (cima) y ***bottom*** (fondo), pero solamente los más ligeros, los *up* y *down*, contribuyen sustancialmente a la formación de protones y neutrones. (Los quarks más pesados desempeñan un pequeño papel como partículas virtuales). El quark *up* tiene carga eléctrica $+^{2}/_{3}$ y el quark *down* tiene carga $-\ ^{1}/_{3}$, y hay tres quarks en cada nucleón. Jugando un poco con esas cargas, puede convencerse de que debe haber dos *ups* y un *down* en cada protón, y dos *downs* y un *up* en cada neutrón. Estaría en lo cierto. Los fermiones que no sienten las interacciones fuertes se conocen como **leptones**, de los que se conocen seis sabores: el electrón, muon y tau, todos con carga –1, y los correspondientes neutrinos: el neutrino electrónico, el neutrino muónico y el neutrino tau, todos con carga 0.

La masa del quark *up* es de unos 2,2 MeV, y la masa del quark *down* es aproximadamente 4,7 MeV. Inmediatamente nos damos cuenta de dos cosas: primero, la masa combinada de los quarks en un protón es mucho menor que la propia masa del protón (938,3 MeV), y lo mismo para el neutrón (939,6 MeV). ¿De dónde procede toda esa masa adicional? En segundo lugar, en efecto, la longitud de onda

Compton de cada quark es mucho mayor que la longitud de onda Compton de los propios nucleones. ¿Cómo encajan estos datos?

Ambas cuestiones surgen porque el interior de un nucleón es un lugar en el que nuestra intuición, entrenada como está en partículas y mecánica clásica, nos falla por completo. Ni siquiera nuestras fantasiosas formas de pensar con diagramas de Feynman van a estar a la altura. Las interacciones entre quarks y gluones son fuertes, no débiles. Encontramos que la TCC predice partículas observando en primer lugar una teoría de campo libre (sin interacciones) e introduciendo interacciones entre las partículas mediante la teoría de perturbaciones. Dentro de un nucleón no hay ningún parámetro pequeño, como la constante de estructura fina, que haga posible la teoría de perturbaciones, y no podemos pensar con seguridad en lo que ocurre en términos de una colección de partículas que interaccionan débilmente. Los nucleones son en su integridad objetos de teoría cuántica de campos y necesitamos tomarnos en serio la TCC para entenderlos.

El hecho de que las interacciones fuertes sean realmente fuertes nos lleva a la característica crucial que ayuda a responder a nuestras preguntas: el **confinamiento**. La carga eléctrica es una variable única que puede ser positiva, negativa o cero. A riesgo de sonar innecesariamente difícil de entender, pensemos que la carga eléctrica toma valores en un espacio vectorial unidimensional. De acuerdo con la teoría de la **cromodinámica cuántica** (CDC), los quarks tienen un tipo adicional de **carga de color** que toma valores en un espacio vectorial *tridimensional*. Esto, por supuesto, no tiene absolutamente nada que ver con el «color» tal y como lo percibimos con nuestros ojos, es solamente un nombre caprichoso.[1] Pero sirve para un propósito útil. Si etiquetamos los tres ejes del espacio de carga de color como «rojo», «verde» y «azul», el principio de confinamiento dice que solo las combinaciones incoloras («blancas») forman partículas observables. Los quarks van a ser individualmente rojos, verdes o azules, y nunca vamos a verlos solos en la naturaleza. Lo que vemos

son combinaciones de tres quarks, cada uno de un color primario distinto, que se suman para dar un protón o un neutrón incoloro.

Piénselo así. Un único átomo de hidrógeno contiene un electrón y un protón. Pero su masa es ligeramente inferior a la suma de esas dos partículas, en unos 13,6 eV. Esto se debe a que tanto el electrón como el protón, en sí mismos, están rodeados de campos eléctricos, y estos campos transportan energía. En un átomo, esos campos (en gran medida) se anulan entre sí, por lo que esa energía no está ahí. Como resultado, la energía total (y, por tanto, la masa) es un poco menor. La discrepancia de 13,6 eV es la **energía de enlace** del electrón en el hidrógeno. Hay que aportar, al menos, esa cantidad de energía para separar las dos partículas.

Consideremos ahora un quark solo, no dentro de un protón. No existe tal cosa, pero imaginemos que existiera. Estaría rodeado por un campo cromodinámico, el campo que da lugar a los gluones de la misma manera que un campo electromagnético da lugar a los fotones. Pero, mientras que el campo electromagnético decae como la inversa del cuadrado (ley de Coulomb), el campo cromodinámico se hace cada vez más grande con la distancia al quark. En consecuencia, tendría que haber una cantidad infinita de energía en ese campo de gluones. En cambio, si tomamos tres quarks en una combinación incolora, el campo de gluones se anula a partir de cierta distancia, al igual que ocurre con el campo eléctrico en un átomo. Lo que queda es un nucleón de energía finita, en el que las energías de enlace de los quarks son efectivamente infinitas.

Además de los protones y los neutrones, existen otras combinaciones incoloras de tres quarks: partículas que implican diferentes sabores de quarks, o estados de mayor energía de los tres quarks. En conjunto, estas partículas se denominan **bariones**. Hasta donde han revelado los experimentos actuales, el número total de bariones menos el número total de antibariones en un proceso —el **número de bariones** neto— se conserva. Se puede pensar igualmente en esto como una conservación del «número de quarks» neto, pero ya

conocíamos la conservación del número de bariones antes de saber que los bariones estaban formados por quarks, así que esta es la etiqueta que usamos. Sin embargo, el número total de bariones en todo el universo es mayor que el número de antibariones. Esta asimetría materia/antimateria aún no se ha explicado según la física conocida. (El **número de Lepton**, que es el número de leptones menos el de antileptones, se conserva por separado. Pero no sabemos si hay un exceso de leptones sobre antileptones, ya que no conocemos el número total de neutrinos y antineutrinos en el universo).

Hay otra forma de hacer una combinación incolora: un quark y un antiquark. Los antiquarks son *antirrojos*, *antiazules* o *antiverdes*, que pueden considerarse como cian, amarillo y magenta. Una combinación quark/antiquark se denomina **mesón**, en contraste con los bariones de tres quarks. Es de esperar que los mesones se desintegren rápidamente, como suele ocurrir, pero, dependiendo del tipo de quarks que los compongan, pueden permanecer en el espacio el tiempo suficiente para ser detectados.

Ahora estamos en condiciones de responder a nuestras dos preguntas. ¿Por qué los protones y los neutrones son mucho más masivos que los quarks que los componen, y cómo encajan esos quarks de baja masa dentro de la longitud de onda Compton de un nucleón? Las respuestas se reducen al hecho de que pensar en un nucleón como si fueran simplemente tres partículas de quarks orbitando entre sí, es una imagen terriblemente engañosa. En realidad, se trata de complicadas configuraciones cuánticas de campos de quarks y gluones. La afirmación honesta es simplemente que el protón corresponde al estado de menor energía, con un número de bariones igual a 1. La masa total de un nucleón procede de los campos de gluones incluso más que de los campos de quarks; si los quarks tuvieran masa cero, las masas de protones y neutrones no serían muy diferentes. Y no debería sorprender que los quarks puedan comprimirse en una región más pequeña que su longitud de onda Compton, a pesar de que esa longitud de onda representa el tamaño

más pequeño para una sola partícula en sí misma. Los quarks dentro de los bariones *no* están solos; se han establecido en un estado cuántico particular en constante interacción con otros campos de quarks y con los gluones.

A veces se oye describir el interior de un protón como un lugar salvaje, con gluones fluctuantes y pares quark/antiquark que surgen y dejan de existir. Incluso los físicos profesionales hablan de este modo. Pero ya hemos aprendido suficiente teoría cuántica de campos como para saber que no es así. Una imagen así es tan engañosa como pensar que un electrón orbita alrededor de un núcleo como un pequeño planeta en un sistema solar. La realidad es una función de onda cuántica. La función de onda de los campos en un protón es una combinación específica y altamente entrelazada de campos de quarks y gluones. Ni siquiera es exactamente correcto, aunque en la práctica es perdonable, decir «hay tres quarks en un protón». Hay tres quarks que contribuyen al número de bariones del protón, conocidos como **quarks de valencia**, pero la configuración fundamentalmente no tiene carácter de partícula. Y hay otros campos que contribuyen a lo que ocurre, como las partículas virtuales en un diagrama de Feynman.

La configuración del campo en un nucleón también es completamente estática —la función de onda no cambia de un momento a otro en modo alguno. Si de alguna manera pudiéramos realizar una medición cuántica del estado de un quark o un gluon dentro de un nucleón, «colapsaría» en algún lugar concreto, y las mediciones repetidas colapsarían en estados diferentes, dando la impresión de que son «fluctuaciones aleatorias». Pero para la mayoría de los protones nadie mide en su interior y nada fluctúa en absoluto. Se trata de soluciones estáticas de la ecuación de Schrödinger para el Modelo Estándar con un número de bariones igual a 1.

RADIO DE BOHR

Esto ha sido intenso. Demos un salto hacia arriba en la escala hasta los simpáticos confines del átomo de hidrógeno.

Un átomo de hidrógeno básico está formado por un protón y un electrón. En unidades cotidianas, esas partículas tienen longitudes de onda Compton (reducidas)

$$\lambda_p = 2{,}1\times10^{-14}\ \text{cm}\ , \qquad \lambda_e = 3{,}9\times10^{-11}\ \text{cm}. \tag{7.11}$$

La longitud de onda del electrón es obviamente mayor, por lo que es el electrón el que establece el tamaño del átomo en sí; el protón solamente se sitúa localizado en el centro. Por eso los electrones hacen todo el trabajo de la física atómica y molecular, por no hablar de la química y la biología.

Ahora tratemos de averiguar el tamaño de un átomo de hidrógeno real. No es una cifra completamente bien definida, ya que estamos hablando de una función de onda cuántica continua, no de un objeto sólido con un límite preciso. Pero una buena medida aproximada viene dada por el **radio de Bohr**, definido como la distancia más probable desde el protón a la que podríamos observar un electrón si la midiéramos. Viene dado por

$$a_0 = 5{,}3\times10^{-9}\ \text{cm}. \tag{7.12}$$

Es evidente que no es igual a la longitud de onda Compton del electrón; es más de cien veces mayor. ¿Qué está pasando?

La longitud de onda Compton es el tamaño mínimo al que puede localizarse una partícula, pero ciertamente es posible extender la función de onda de esa partícula en una distancia mayor. En el caso del átomo de hidrógeno, tenemos que considerar la fuerza eléctrica del protón que mantiene al electrón unido a él. Podemos pensar en ella en términos de energía potencial en función de la distancia:

$$V(r) = -\frac{\alpha}{r}, \tag{7.13}$$

donde $\alpha \approx {}^{1}/_{137}$ es la constante de estructura fina. Es esta constante de estructura fina la que nos indica la fuerza del electromagnetismo; el hecho de que sea pequeña, pero no demasiado pequeña, significa que la fuerza de enlace eléctrico es débil, pero no demasiado débil.

Lo que sucede, entonces, es que el electrón se dispersa un poco en el átomo de hidrógeno, porque está enlazado por el potencial electromagnético débil, pero no demasiado débil. Si introducimos los números, encontramos

$$a_0 = \frac{\lambda_e}{\alpha}. \tag{7.14}$$

Esto debería ser lo menos sorprendente del mundo, porque se escala de la manera correcta. La escala de longitud para un electrón en un átomo de hidrógeno está «fijada» por la longitud de onda Compton del electrón, porque es el parámetro relevante que tiene dimensiones de longitud. Pero está modificada por el parámetro adimensional α, que caracteriza la fuerza de la interacción electromagnética. Y tiene sentido que sea inversamente proporcional a α: si imaginamos que $\alpha \to 0$, haciendo electromagnetismo más débil, el electrón estaría cada vez menos ligado al protón y acabaría dispersándose por todo el espacio. En cambio, si hiciéramos α más grande, el electrón se acercaría más al protón.

Se trata de un ejemplo sencillo de **argumento de escala**: se parte de escalas de longitud/tiempo/energía básicas (como la longitud de onda Compton del electrón) y se utiliza la intuición física para averiguar cómo se ven afectadas por otros parámetros (como la constante de estructura fina) para estimar algunas cantidades de interés (como el radio de Bohr). Observe que hemos sido capaces de averiguar esto sin llegar siquiera a resolver realmente la ecuación de Schrödinger con

nivel de detalle alguno, solamente mediante escalado y análisis dimensional. Esta es la razón por la que los físicos siempre piensan primero en ese tipo de argumentos antes de hacer ningún trabajo duro. (Como advertía el físico John Wheeler, «Nunca hagas un cálculo hasta que sepas la respuesta»).

Incluso podemos ir un poco más lejos, para comprender el origen de los 13,6 eV de energía de enlace del electrón en el hidrógeno. Mediante un cálculo clásico aproximado (que funciona bastante bien aquí), la energía del electrón a una distancia determinada es solamente su energía potencial (7.13). Y la distancia típica es el radio de Bohr (7.14). Combinando estas dos ecuaciones, llegamos a la conclusión de que la energía de enlace debería ser

$$E_1 \approx \frac{\alpha^2}{\lambda_e} = \alpha^2 m_e = 27{,}2 \text{ eV}. \qquad (7.15)$$

¡Oh, no, hemos obtenido una respuesta incorrecta! No se preocupe. Si se fija bien, en realidad tenemos el doble de la respuesta correcta. Eso no es un fallo, es un resultado bastante bueno, teniendo en cuenta que solamente hemos utilizado el análisis dimensional y algo de gesticulación. Esta técnica no sería suficientemente buena para hacer física de precisión, pero nos lleva por buen camino y nos da una idea aproximada de cómo dependen unas cantidades de otras.

MOLÉCULAS

La relación de masas entre el protón y el electrón, $m_p/m_e \approx 1800$, podría parecer una jerarquía que pide explicaciones. Pero no es un número muy grande, y al fin y al cabo alguna partícula tiene que ser la más ligera. Del mismo modo, $\alpha \approx {}^1/_{137}$ es pequeño, pero no exageradamente pequeño. Quizá simplemente las cosas sean así.

Sin embargo, una vez que tenemos esos números en la mano, podemos empezar a entender la cascada de escalas de energía que nos lleva de las partículas elementales al mundo cotidiano. La energía de enlace de un electrón en el hidrógeno es del orden de 10^1 eV, menor que la masa del electrón en un factor de α^2. En átomos más masivos, los electrones más externos están generalmente más alejados de sus núcleos, y sus energías de enlace son aún menores. Por tanto, los números típicos se sitúan entre 1 y 10 eV. Por razones similares, esas son también las escalas que caracterizan la energía de enlace (o disociación) de los átomos entre sí, dentro de las moléculas. Es decir, esas son las energías de las reacciones químicas. La química surge de átomos que comparten (o dejan de compartir) electrones entre sí, todo lo cual, en última instancia, se reduce a la masa del electrón y a potencias apropiadas de la constante de estructura fina.

También hay energías asociadas a las propias moléculas. Una molécula puede vibrar cuando dos núcleos atómicos de su interior se mueven acercándose y alejándose. Únicamente por la descripción, no le sorprenderá saber que existen niveles de energía para esos movimientos, que son bien aproximados por nuestro amigo el oscilador armónico simple. Las escalas de energía típicas de las vibraciones moleculares son del orden de 10^{-2} eV. También hay tipos más sutiles de movimientos moleculares, en los que la forma general de la molécula se retuerce o distorsiona. Estos son algo más suaves, con energías típicamente del orden de 10^{-3} eV.

Y así sucesivamente a medida que consideramos moléculas más complicadas, como las que encontramos en la química orgánica. La existencia de un gran número de formas de vibración en una molécula significa que los ejemplos de menor energía probablemente sean incluso de energía inferior a la de las rotaciones típicas de una molécula simple. Una molécula extensa y complicada tendrá vibraciones con longitudes de onda y escalas de tiempo relativamente largas, que corresponden a energías pequeñas. De este modo, la química construye una rica jerarquía de escalas de energía, todas ellas

muy inferiores a las que estamos acostumbrados a tratar en la física de partículas. Nunca dejará de sorprendernos la complejidad de nuestro mundo cotidiano, construido a partir de unos pocos ingredientes elementales.

---

1 Los «sabores» de los quarks tampoco son sabores que se puedan degustar. Siento decepcionarle.

OCHO

SIMETRÍA

El primer capítulo de *Espacio, tiempo y movimiento* estaba dedicado a la idea de «conservación». Una de las cosas que mencionamos fue el teorema de Noether, según el cual toda transformación continua de simetría de un sistema está asociada a una cantidad conservada. Y dijimos que una simetría es una transformación que se puede hacer a un sistema dejando sus características esenciales inalteradas.

Todo es cierto. Y las simetrías son útiles en toda la mecánica clásica, especialmente en la relatividad, donde el hecho de que las leyes de la física no cambien de forma cuando hacemos una transformación de Lorentz es un ejemplo especialmente significativo de simetría en acción. Pero es en la teoría cuántica de campos donde las simetrías pasan de ser «realmente útiles» a «absolutamente indispensables». En cierto sentido, las fuerzas de la naturaleza surgen directamente de un tipo particular de simetría de los campos subyacentes, conocida como «invariancia de gauge». Dedicaremos un capítulo entero a ello, pero primero tenemos que entender la idea general de simetría.

Y eso significa que tenemos que pensar en la **teoría de grupos**, que es el marco matemático para clasificar y utilizar las simetrías. Los conceptos que se manejan son bastante sencillos —ciertamente, si se comparan con el cálculo, la geometría no euclidiana o, de hecho, la mecánica cuántica—, pero el lenguaje y la notación pueden parecer abstractos y poco familiares, por lo que merece la pena ir poco a poco. Para no perder el norte, he aquí un breve resumen de lo que veremos en este capítulo:

— Las simetrías son transformaciones que podemos hacer en las cosas —ya sean formas geométricas o conjuntos de campos cuánticos— que dejan inalterada su esencia.
— Algunos ejemplos sencillos son cómo podemos girar o voltear formas como triángulos y cuadrados. Estas son simetrías «discretas», ya que existe un conjunto discreto de transformaciones que podemos realizar.
— Los conjuntos de simetrías se describen matemáticamente como «grupos». Si pensamos en las propiedades de los grupos en general, podemos comprender mejor cómo funcionan las simetrías en física.
— Existen simetrías tanto continuas como discretas, como el conjunto de rotaciones en un determinado número de dimensiones. Las rotaciones en un espacio de n dimensiones se describen mediante los grupos ortogonales SO(n).
— También existen espacios de n dimensiones complejas, útiles para describir partículas cargadas. En ese caso, las rotaciones se describen por los grupos unitarios SU(n). Hay muchos detalles matemáticos asociados a los grupos unitarios, pero la idea básica es simplemente que describen «rotaciones de vectores hechas a partir de números complejos».

La razón por la que el Modelo Estándar de la física de partículas se conoce a veces como «SU(3)×SU(2)×U(1)» es porque esa es la etiqueta del conjunto específico de simetrías de gauge que aparecen (hasta donde sabemos) en el mundo real. Empecemos a averiguar qué significa eso.

INVARIANCIA

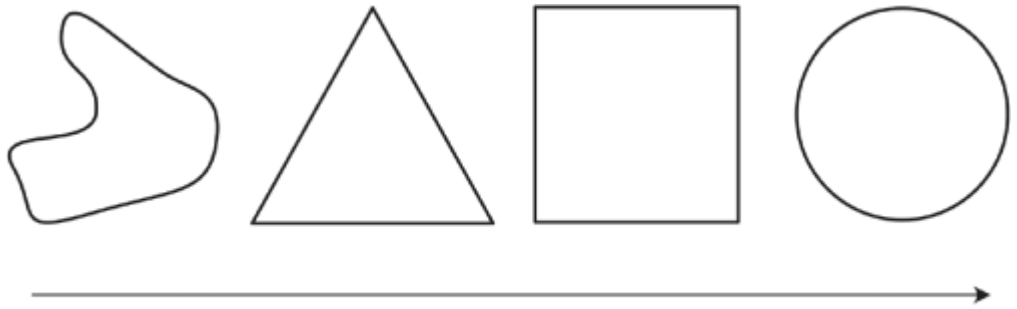
Los seres humanos tendemos a ver el mundo en términos humanos. Nuestras ideas sobre la simetría, por ejemplo, están influidas por cómo

miramos a otros seres humanos. Según algunas pruebas antropológicas, tendemos a juzgar los rasgos faciales simétricos como generalmente más atractivos. En este caso, la simetría en cuestión es bilateral: imaginamos un rostro reflejado en torno a una línea vertical que pasa por el centro y lo juzgamos simétrico si los dos lados son casi imágenes especulares el uno del otro. (Un rostro que fuera simétrico con respecto a una línea horizontal trazada en el centro parecería horrendo). O, más informalmente, pensamos en la simetría como un aspecto del equilibrio o la proporción general de una persona o un objeto.

Pero la simetría bilateral es solamente un ejemplo de un concepto mucho más general. En física o matemáticas a menudo consideramos la idea de **transformación** de un objeto: algún proceso (matemáticamente, un mapeado o función) que modifica el objeto de alguna manera, ya sea intrínsecamente o en relación con el resto del mundo. Tomar y mover un objeto es una transformación, pero también lo es distorsionar su configuración espacial o reflejarlo respecto a algún plano. No es necesario que realicemos físicamente una transformación concreta para preguntarnos cuáles serían sus propiedades si la hiciéramos.

Si una transformación deja inalterada alguna propiedad, decimos que la propiedad es **invariante** bajo la transformación. Y cuando existe tal invariancia, la transformación correspondiente es oficialmente una **simetría**. En última instancia, nos interesarán las simetrías que se obtienen transformando unos campos cuánticos en otros. Las interacciones fuertes, por ejemplo, se basan en una simetría denominada SU(3) que rota los tres colores de los quarks entre sí. Pero podemos empezar de forma más concreta pensando en las simetrías de las formas geométricas.

Consideremos las siguientes figuras bidimensionales: una curva amorfa cerrada, un triángulo equilátero, un cuadrado y un círculo.



Las formas se han mencionado en orden de simetría creciente. La curva amorfa es la menos simétrica, seguida del triángulo, el cuadrado y, por último, el círculo.

Para la curva amorfa, esto tiene mucho sentido: no hay mucha simetría en ella. Del mismo modo, el círculo parece claramente el más simétrico. Es difícil ser más simétrico que un círculo perfecto.

En los casos del triángulo y el cuadrado, la situación es menos clara. Ambos parecen bastante simétricos, quizá por igual. Pero hay un sentido riguroso en el que el cuadrado tiene más simetría que el triángulo. Para verlo, tendremos que ser más precisos sobre cómo se clasifican matemáticamente las simetrías.

SIMETRÍAS TRIANGULARES

Consideremos el triángulo con cierto detalle. ¿Cuánta simetría tiene, y qué significa incluso preguntar «cuánta simetría» hay en una situación concreta?

Una simetría es una transformación que deja invariante un objeto. Así que tenemos que ser muy cuidadosos con el «objeto» que estamos considerando y con el tipo de transformaciones que nos interesan. En el caso del triángulo, vamos a etiquetar sus vértices para saber qué le estamos haciendo: el vértice superior es A y, moviéndonos en el sentido de las agujas del reloj, tenemos B en el vértice inferior derecho y C en el inferior izquierdo. «El triángulo» es la figura geometría pura

en sí, no las etiquetas; estas solamente nos ayudan a seguir la pista de cómo lo estamos transformando.

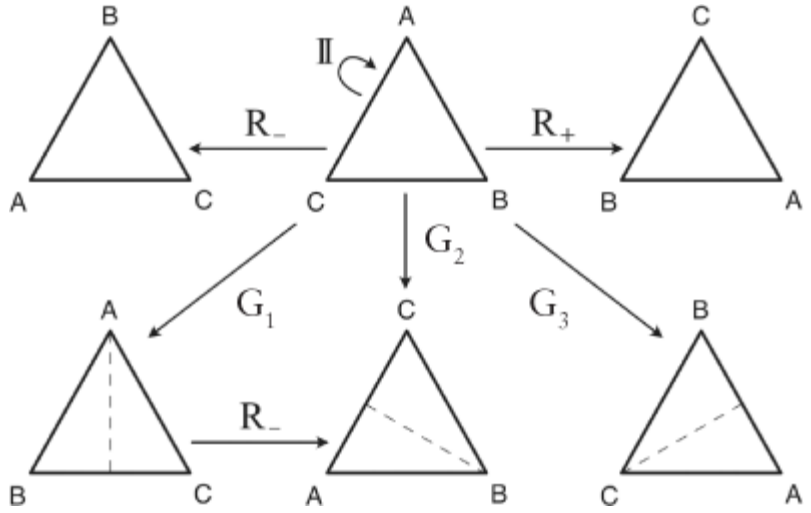
Si rotamos el triángulo 120 grados en el sentido de las agujas del reloj (un tercio de una rotación completa), el vértice A se desplaza a donde estaba el B, el B se desplaza a donde estaba el C, y el C se mueve a donde estaba el A. La forma resultante es idéntica a la forma con la que empezamos. Así que esta **rotación** de 120 grados en el sentido de las agujas del reloj es una transformación, etiquetada R_+, y esa transformación es de hecho una simetría. (Usamos el signo + para indicar «en el sentido de las agujas del reloj»). Rotar en el sentido de las agujas del reloj 90 grados también es una transformación N_+ que podemos hacer sobre el triángulo, pero no es una simetría, ya que el triángulo obviamente se ve diferente después de hacerlo.

Ahora podemos preguntarnos, ¿cuáles son *todas* las simetrías del triángulo? Está claro que podemos rotarlo 120 grados en el sentido de las agujas del reloj. No nos sorprenderá que también podamos rotarlo 240 grados en el sentido de las agujas del reloj (dos tercios de una rotación completa). O, para el caso, 360 grados (una rotación completa). O cualquier otro múltiplo de 120 grados. Además, también podríamos rotar en la otra dirección (en sentido contrario a las agujas del reloj) 120 grados; llamemos a esa transformación R_-. Y podemos rotar en sentido contrario a las agujas del reloj en cualquier múltiplo de 120 grados. Todas estas transformaciones son simetrías, ya que dejan invariante la forma original.

También hay simetrías del triángulo que no son rotaciones en absoluto. Considere la reflexión alrededor de una línea vertical, trazada desde el vértice superior y bisecando el triángulo. Este **giro**,

que podemos llamar G_1, tiene el efecto de intercambiar los dos vértices inferiores sin cambiar el superior. (También podría llamarse «reflexión», pero la letra R ya está reservada para las rotaciones). Del mismo modo, podemos considerar un giro G_2 alrededor de una línea bisectriz trazada desde el vértice inferior derecho, y G_3, un giro alrededor de una línea bisectriz trazada desde el vértice inferior izquierdo.

Al intercambiar solamente dos de los vértices, cualquiera de las transformaciones de giro cambia la **paridad** o «lateralidad» del triángulo. Partiendo de A en el sentido de las agujas del reloj, llegaríamos primeramente a B y luego a C, y este hecho no cambia al realizar las rotaciones. Pero si realizamos cualquiera de las operaciones de giro, si partimos de A en el sentido de las agujas del reloj llegaremos a C y después a B. Así es como podemos estar seguros de que ninguna combinación de rotaciones es equivalente a un giro. (Cuando estamos confinados a dos dimensiones en cualquier caso). Todas las transformaciones que hemos considerado se representan en la figura, partiendo del triángulo original en la parte superior central.



Sin duda se habrá dado cuenta de que estas transformaciones no son independientes entre sí: una transformación puede ser equivalente a una combinación de otras. Por ejemplo, hacer R_+ dos veces (es decir,

rotar 240 grados en el sentido de las agujas del reloj) equivale a hacer R_- (rotar 120 grados en el sentido contrario a las agujas del reloj) solamente una vez. En cualquiera de los dos casos, acabamos con el vértice A abajo a la izquierda, B arriba y C abajo a la derecha. El triángulo no recuerda, ni le importa, en qué dirección se movió; lo único que le importa es la configuración final de la forma después de hacer la transformación. Por tanto, rotar 360 grados en cualquier dirección equivale a no hacer nada. Tenemos que tener en cuenta cada transformación de simetría posible, incluso una tan trivial —«no hacer nada en absoluto»— que se denomina **transformación de identidad**, o simplemente $\mathbb{I}$. Y como podemos ver en la figura, hacer un giro G_1 seguido de una rotación R_- es equivalente a hacer un giro G_2.

GRUPOS DE SIMETRÍAS

Puede juguetear con otras combinaciones de nuestras transformaciones de simetría y ver qué obtiene. Mejor aún: convénzase de que no nos falta ninguna simetría. Todo lo que pueda hacer a un triángulo equilátero y que le haga parecer exactamente el mismo triángulo equilátero puede considerarse una rotación, un giro o la identidad (manteniendo inalterado el triángulo original). Este conjunto completo se denomina **grupo** de simetrías. En el caso del triángulo equilátero, el grupo de simetrías se conoce como «grupo diédrico de grado 3», que suena mucho más intimidatorio de lo que es en realidad nuestra colección de rotaciones y giros. Se escribe D_3, y está formado por los seis elementos

$$D_3 = \{\mathbb{I}, R_+, R_-, G_1, G_2, G_3\}. \qquad (8.1)$$

Nadie debería molestarse si lo llama simplemente «el grupo del triángulo».

La razón es que el grupo diédrico de «grado 3» deriva de que el triángulo tiene tres vértices, que se intercambian entre sí bajo la acción del grupo. El grupo de simetrías de un cuadrado sería D_4, para un octógono sería D_8, y así sucesivamente. Para cualquier polígono regular con n vértices, el grupo de simetrías Dn tiene 2n elementos: n rotaciones (incluida la identidad) y n giros. Por eso decimos que el cuadrado ($n = 4$) tiene más simetría que el triángulo: hay ocho elementos en su grupo de simetrías, frente a los solamente seis del triángulo. Puede resolverlos jugando con rotaciones y reflexiones que dejen invariante el cuadrado. Pista: hay giros alrededor de ejes que van de un vértice al opuesto, y también giros que van del centro de un lado al opuesto.

Nuestro triángulo es una cosa bastante simple. Probablemente pueda visualizar mentalmente lo que ocurre cuando lo giramos o rotamos. Ni que decir tiene que habrá simetrías más complicadas y sutiles para las que la visualización nos fallará. Cuando nos encontramos por primera vez con ideas complicadas en física o matemáticas, nos solemos preguntar: «¿Cómo se visualiza eso?». La respuesta corta es: «No se visualiza». La respuesta más larga es: «Puede intentar visualizar análogos simplificados de menos dimensiones de las que estamos hablando, pero en última instancia está construyendo una estructura matemática e intentando averiguar qué implican las ecuaciones, en lugar de visualizar directamente lo que tiene en mente». (Lo mismo puede decirse de la mecánica cuántica y la teoría de campos en sentido más general). La buena noticia es que las matemáticas son bastante fiables, si se utilizan correctamente.

Estaría bien disponer de alguna notación para combinar simetrías, por ejemplo, para indicar que un giro G_1 seguido de una rotación R_- equivale a hacer un giro G_2. Así es como vamos a escribir ese hecho:

$$R_- \cdot G_1 = G_2. \qquad (8.2)$$

Aquí están sucediendo un par de cosas. A la izquierda parece que estamos multiplicando R_- por G_1 y, de hecho, los matemáticos a veces lo expresan de ese modo. Pero esta multiplicación no es exactamente igual que la idea habitual que tenemos en mente cuando multiplicamos números. Con «$R_- \cdot G_1$» queremos decir «primero hacemos el giro G_1, y luego la rotación R_-». Eso equivale a leerlo de derecha a izquierda, lo que podría parecer que está al revés, pero en nuestra mente estamos pensando en estas operaciones como transformaciones que actúan sobre el triángulo. Al pensar en nuestras transformaciones como funciones, se pueden entender (8.2) como la expresión $R_-(G_1(\Delta)) = G_2(\Delta)$, donde Δ indica el triángulo considerado.

Y lo que es más importante, estas transformaciones no **conmutan**, lo que es una forma de decir que el orden en que escribimos las transformaciones importa. Volviendo a la figura, se puede deducir que haciendo estas transformaciones en el orden inverso, R_- y luego G_1, nos daría un triángulo con B en la parte superior, A en la parte inferior derecha, y C en la parte inferior izquierda. Eso equivale a aplicar G_3 sobre el triángulo original, en lugar de G_2. Así que escribiríamos $G_1 \cdot R_- = G_3$. Combinando todos nuestros elementos del grupo de varias maneras, podríamos elaborar una «tabla de multiplicar» (aunque no sea la multiplicación habitual) que resumiría toda la estructura del grupo.

Aquí hemos escrito explícitamente la tabla de multiplicar del grupo para D_3. La convención es que en la operación del grupo, el elemento que encabeza la columna va a la derecha (y, por lo tanto, actúa primero) y el elemento que encabeza la fila va a la izquierda. Si observamos la columna G_1 y la fila R_-, tenemos $R_- \cdot G_1 = G_2$, como era de esperar.

D_3	$\mathbb{I}$	R_+	R_-	G_1	G_2	G_3
$\mathbb{I}$	$\mathbb{I}$	R_+	R_-	G_1	G_2	G_3
R_+	R_+	R_-	$\mathbb{I}$	G_3	G_1	G_2
R_-	R_-	$\mathbb{I}$	R_+	G_2	G_3	G_1
G_1	G_1	G_2	G_3	$\mathbb{I}$	R_+	R_-
G_2	G_2	G_3	G_1	R_-	$\mathbb{I}$	R_+
G_3	G_3	G_1	G_2	R_+	R_-	$\mathbb{I}$

Una vez que nos convertimos en matemáticos de élite, tendemos a dejar de pensar en los grupos como «conjuntos de transformaciones de simetría sobre objetos» y empezamos a pensar en ellos como «conjuntos de cosas que podemos combinar entre sí en un orden especificado para obtener otro elemento del conjunto». Hay algunas restricciones, o **axiomas de grupo**, sobre cuándo un conjunto de este tipo puede considerarse un grupo, principalmente, que un único elemento debe ser la identidad y que cada elemento tiene un inverso. Decir que «existe un elemento de identidad $\mathbb{I}$» significa que, para cualquier elemento de grupo g, ese elemento no se modifica al actuar con $\mathbb{I}$ a la derecha o a la izquierda:

$$g \cdot \mathbb{I} = \mathbb{I} \cdot g = g. \tag{8.3}$$

Y «todo elemento tiene un inverso» significa que, dado g, existe un elemento g^{-1} que da la identidad cuando lo multiplicamos por g por cualquier lado:

$$g \cdot g^{-1} = g^{-1} \cdot g = \mathbb{I}. \tag{8.4}$$

Expresándolo con palabras: en cada grupo siempre hay una transformación que deja el objeto inalterado, y cualquier transformación en el grupo puede ser deshecha por alguna otra transformación.[1] El inverso de un elemento puede ser el propio elemento; eso es cierto, por ejemplo, para cada uno de nuestros giros en el grupo del triángulo.

TEORÍA DE GRUPOS

A los ojos de un matemático, un grupo no es más que un conjunto con un tipo particular de estructura: básicamente, una «regla de multiplicación» que tiene un elemento de identidad y un inverso. Esta idea desempeña un papel especial en el estudio de las simetrías, porque en el conjunto de las transformaciones de simetría hay de hecho siempre un elemento de identidad (no hacer nada) y un elemento inverso a cada transformación (deshacer lo que se hizo originalmente).

Uno de los obstáculos para aprender teoría de grupos por primera vez es que las ideas de «multiplicación» y «tener un inverso» nos suenan familiares de cómo manipulamos los números ordinarios, pero hay distinciones cruciales que no se pueden ignorar. Consideremos el conjunto de los números reales y, para nuestra regla de multiplicación, tomemos la multiplicación a la vieja usanza: $2 \times 3 = 6$, ese tipo de cosas. Inmediatamente nos damos cuenta de que *no* se trata de un grupo, porque el requisito de que cada elemento tenga un inverso no se cumple: 0 no tiene inverso. Existe un elemento de identidad, el 1, pero eso no basta para salvarlo. Para ser un grupo, tiene que satisfacer todos los axiomas de grupo sin excepción.

Cuando pensamos en un posible grupo, es importante especificar no solamente los elementos del conjunto, sino también la operación de «multiplicación» binaria específica que tenemos en mente. Los números reales bajo multiplicación ordinaria no son un grupo, pero pensemos en los **números reales con la suma**: $2 + 3 = 5$ y así

sucesivamente. *Hay* un grupo, que denominamos $\mathbb{R}$, y al que a veces llamamos perezosamente «los números reales» o, simplemente, «los reales». En este grupo, el inverso de x es simplemente $-x$, y el elemento de identidad es 0. La «regla de multiplicación» es lo que crecimos pensando que era la suma. Tendremos que acostumbrarnos.

Ahora ya no hay nada que nos detenga, al menos en lo relativo a la teoría de grupos. Dados los axiomas que hemos especificado para definir un grupo, podemos empezar a hacernos todo tipo de preguntas sobre cómo funcionan los grupos y cómo se relacionan entre sí los distintos grupos. Por ejemplo, se dice que dos grupos son **isomorfos** entre sí si son realmente el «mismo» grupo, aunque sus definiciones parezcan diferentes. Un isomorfismo entre dos grupos es una correspondencia de uno a uno entre los elementos de los dos conjuntos, que tiene la propiedad de que la regla de multiplicación se conserva mediante esta correspondencia. Es decir, multiplicar en un grupo y luego mapear los resultados al otro nos da la misma respuesta que mapear de un grupo al otro y después multiplicar allí. Si dos grupos G y H son isomorfos, lo escribimos como $G \cong H$.

También tenemos la noción crucial de **subgrupo**: un subconjunto del conjunto de elementos del grupo que forma un grupo por sí mismo, con respecto a la misma regla de multiplicación. En la práctica, esto equivale a comprobar la propiedad de clausura: en un subgrupo, tenemos que poder combinar dos elementos cualesquiera y seguir obteniendo un elemento del subgrupo, sin saltarnos al grupo mayor. En el grupo del triángulo, el conjunto de rotaciones $\{\mathbb{I}, R_+, R_-\}$ es un subgrupo: la combinación de dos rotaciones cualesquiera sigue siendo una rotación. Pero algo como $\{\mathbb{I}, R_+\}$ no es un subgrupo, ya que $R_+ \cdot R_+ = R_-$, que no está en la lista de elementos de $\{\mathbb{I}, R_+\}$.

NÚMEROS ENTEROS

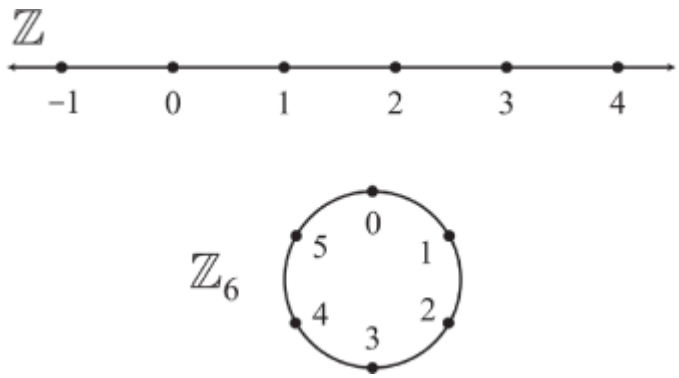
Los números **enteros**, etiquetados $\mathbb{Z}$, son un subgrupo obvio de los reales, $\mathbb{R}$. Los elementos son el conjunto infinito {... –1, –2, 0, 1, 2, 3, ...}, y la suma es, una vez más, la operación binaria apropiada. (Así que, en realidad, deberíamos llamar a este grupo «los enteros con la suma», pero se espera que complete las condiciones implícitas). Es tradicional, para los reales o los enteros o cualquier otro grupo en el que combinamos elementos mediante adición, utilizar el signo + habitual para denotar la operación binaria, en lugar del signo de punto en el centro, ·. Obsérvese que empezamos, en nuestra discusión del grupo del triángulo, pensando en los elementos del grupo como transformaciones de un objeto. Ahora, con los reales o enteros, pensamos en los elementos del grupo como objetos en sí mismos: los números de toda la vida. No pasa nada. En lo que respecta a la teoría de grupos, no son más que elementos de algún conjunto sobre el que definimos alguna operación binaria.

He aquí una pregunta para contemplar: dado que ambos tienen un número infinito de elementos, ¿son $\mathbb{Z}$ y $\mathbb{R}$ isomorfos entre sí? (Pista: no, no lo son. No todos los infinitos se crean de la misma manera).

Al pensar en subgrupos, es importante recordar que los grupos son conjuntos de elementos con una determinada operación binaria. Que tengamos un sub*conjunto* no significa que tengamos un sub*grupo*. Tenemos que comprobar que el subconjunto sea cerrado bajo la operación de grupo. Podríamos pensar, por ejemplo, que los números naturales —todos los números enteros iguales o mayores que cero— serían buenos candidatos para ser subgrupo de los números enteros, pero no funciona del todo. Aunque el 0 podría funcionar como un elemento de identidad con la suma, no hay inversos, ya que no hay números naturales negativos. Así que los números naturales no son un grupo en absoluto y, ni mucho menos, un subgrupo.

Tenemos grupos de tamaño finito llamados los **enteros módulo** n, o $\mathbb{Z}_n$ para abreviar. Estos grupos tienen n elementos, {0, 1, 2, ... n–1}. La operación de grupo es la suma como antes, pero ahora es suma «módulo n», o simplemente «mód n». Esto significa que si p y q están

en $\mathbb{Z}_6$, y sumamos $p + q$, si el resultado es 6 o mayor simplemente restamos 6 a la respuesta. Por lo tanto, la suma mód n siempre nos dará un resultado entre 0 y n–1. Esto es sutilmente diferente de la operación de suma ordinaria en el conjunto de los números enteros, por lo que $\mathbb{Z}_n$ no es un subgrupo de $\mathbb{Z}$. Los elementos de $\mathbb{Z}_n$ pueden considerarse como puntos discretos dispuestos en un círculo: una vez que contamos aumentando hasta n–1, volvemos a 0 para empezar de nuevo.



Pensemos en $\mathbb{Z}_6$, los números enteros módulo 6. Tiene seis elementos: {0, 1, 2, 3, 4, 5}. Ese es el mismo número de elementos que nuestro grupo de simetrías del triángulo D_3. Es natural preguntarse si $\mathbb{Z}_6$ es isomorfo a D_3, pero no lo es. La tabla de multiplicación es diferente, así que no son los mismos grupos. En realidad es bastante fácil ver esto, una vez que recordamos que las transformaciones en el grupo del triángulo no conmutan entre sí. Por ejemplo, hemos observado que $R_- \cdot G_1 \neq G_1 \cdot R_-$. Pero la suma, ya sea en los enteros $\mathbb{Z}$ o en un grupo como $\mathbb{Z}_6$, ciertamente conmuta: $p + q = q + p$ para todo p y q. Así que la operación de grupo no puede funcionar igual en $\mathbb{Z}_6$ que en D_3.

Aquellos grupos donde la operación siempre conmuta, $a \cdot b = b \cdot a$ para todo elemento a y b, se denominan **abelianos**, en honor al matemático noruego Niels Henrik Abel. Los grupos en los que la operación a veces no conmuta son, naturalmente, **no abelianos**. $\mathbb{R}$, $\mathbb{Z}$ y $\mathbb{Z}_n$ son todos grupos abelianos, mientras que D_3 es no abeliano. Hay

importantes ejemplos de ambos tipos de grupos en física, pero los grupos no abelianos son obviamente más complicados y, por tanto, más divertidos.

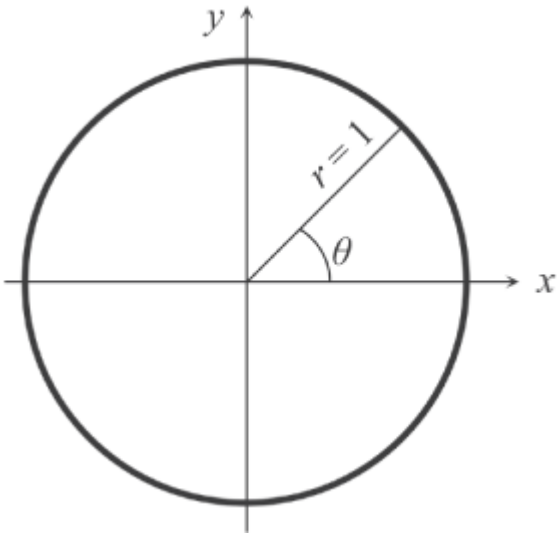
Es importante interiorizar las ideas de que (1) un grupo se define tanto por su conjunto de elementos como por la operación binaria definida sobre ese conjunto, y (2) dos grupos que parecen diferentes son realmente «el mismo», en lo que se refiere a su carácter de grupo, si existe un isomorfismo entre ellos. Consideremos $\mathbb{Z}_2$, los números enteros módulo 2. Como lo hemos definido, los elementos son {0, 1}, y la operación es la suma mód 2, con 0 como el elemento de identidad. Pero también podemos considerar algo llamado «grupo multiplicativo de orden 2», con elementos {+1, −1}, y la operación dada por la multiplicación, de modo que el elemento de identidad es +1. Este grupo es isomorfo a $\mathbb{Z}_2$, y la mayoría de los físicos lo llaman simplemente «$\mathbb{Z}_2$» (para consternación de los matemáticos exigentes). Teniendo en cuenta este isomorfismo, tiene sentido decir que las expresiones «1 + 1 = 0» y «−1 × (−1) = +1» son ambas verdaderas y, de hecho, expresan la misma idea; la primera se formula en el lenguaje de la suma mód 2, mientras que la segunda se formula en el lenguaje de la multiplicación.

$\mathbb{Z}_2$ aparece todo el tiempo en situaciones físicas, ya que a menudo tenemos transformaciones que se deshacen a sí mismas, como tomar la imagen especular de algo o intercambiar partículas con antipartículas. Por ejemplo, el grupo del triángulo tiene subgrupos formados por la identidad y cualquiera de las transformaciones de giro, como $\{\mathbb{I}, G_1\}$. Todos estos subgrupos son isomorfos a $\mathbb{Z}_2$, y la mayoría de los físicos se complacen en referirse a ellos como $\mathbb{Z}_2$, o como «un subgrupo de $\mathbb{Z}_2$». ($\mathbb{Z}_2$ no es un subgrupo de los números enteros, pero es un subgrupo de D_3, porque en este último caso tenemos elementos que dan la identidad cuando se elevan al cuadrado).

SIMETRÍAS DEL CÍRCULO

Hicimos un buen trabajo con el triángulo equilátero, clasificando completamente todas sus transformaciones de simetría, y reduciéndolas al grupo de seis elementos D_3. No sería demasiado difícil hacer un análisis similar del cuadrado y su grupo de simetrías de ocho elementos, D_4. ¿Y con el círculo?

Reconocemos un círculo cuando lo vemos, pero ahora vamos a tener que ser más precisos. La forma más sencilla de definir un círculo es partir de un plano bidimensional, construir coordenadas perpendiculares *x* e *y* sobre él y considerar el conjunto de todos los puntos a una distancia 1 del origen de esas coordenadas. El conjunto de puntos resultante forma un círculo de radio 1. No importa realmente que el punto central sea el origen, o que el radio sea exactamente 1, pero esas opciones nos dan un punto de partida cómodamente definido.



A veces, un círculo se denomina S^1, ya que es un caso especial de la idea más genérica de una esfera *n*-dimensional S_n, definida como «el conjunto de todos los puntos a distancia fija desde el origen en el espacio euclidiano $(n+1)$-dimensional». Nuestra noción habitual de esfera, como la superficie de una pelota de baloncesto, es S^2. Nótese que solamente tenemos en mente la superficie bidimensional de la esfera; el interior tridimensional se denomina *bola*. Existe incluso una

esfera cero, S^0, que es el conjunto de puntos situados a una distancia fija del origen sobre una recta unidimensional. Es decir, solamente los dos puntos en r y $-r$ para el radio r. Este es el tipo de generalización que va más allá de la intuición y que tanto gusta a los matemáticos: una esfera de dimensión cero no es más que un conjunto de dos puntos.

Las simetrías del círculo no son tan diferentes de las del triángulo: hay rotaciones y giros. La diferencia es que los hay en un número infinito. Podemos rotar el círculo cualquier ángulo θ entre 0 y 2π, y del mismo modo podemos reflejar el círculo respecto a cualquier recta que pase por el origen. (Estamos expresando los ángulos en términos de radianes, donde 2π radianes equivalen a 360 grados). Denote-mos una rotación en θ como $R(\theta)$, y una reflexión respecto a un eje en ángulo ϕ por F_ϕ. Entonces el grupo de simetrías del círculo tiene elementos

$$O(2) = \{R(\theta), F_\phi\}. \qquad (8.5)$$

La notación O(2) representa el **grupo ortogonal** en dos dimensiones, que proviene de pensar en el espacio en el que está inmerso el círculo más que en el círculo en sí. Las dos dimensiones en cuestión son los ejes x e y del plano bidimensional. Esos dos ejes son ortogonales (perpendiculares) entre sí, y esa propiedad de ortogonalidad se conserva siempre que rotemos o giremos el plano manteniendo fijo su origen, haya o no un círculo en él. «El conjunto de transformaciones de un plano que mantiene los ejes perpendiculares y el origen fijo» es exactamente el mismo que «el conjunto de simetrías del círculo», y afirmaciones análogas seguirán siendo ciertas en más de dos dimensiones. Por lo tanto, estos grupos se denominan «grupos ortogonales» en lugar de «grupos de esferas», aunque este último término también habría funcionado.

El hecho de que los elementos de O(2) puedan dividirse en rotaciones y giros es importante, pero también entorpece un poco. A menudo conviene concentrarse solamente en el subgrupo que

representa las rotaciones. Este se conoce como el **grupo ortogonal especial** en dos dimensiones, SO(2), con elementos

$$SO(2) = \{R(\theta)\}. \tag{8.6}$$

Si lo piensa, este grupo es en sí mismo un círculo: las rotaciones para cualquier ángulo θ entre 0 y 2π describen un conjunto unidimensional de objetos que topológicamente vuelven a su punto de partida, $R(2\pi) = R(0)$. Pero, en general, no va a ser cierto que el conjunto de rotaciones en n dimensiones forme una esfera de $(n-1)$ dimensiones, se trata solamente de una coincidencia que tenemos cuando $n = 2$.

Esto nos lleva a un punto importante. Cuando hablamos de la relatividad general en *Espacio, tiempo y movimiento*, introdujimos la noción de **variedad**, que es un tipo de espacio que parece un espacio n-dimensional continuo ordinario cuando nos acercamos lo suficiente, aunque podría tener una geometría o topología complicada cuando volvemos a alejarnos. Los grupos O(2) y SO(2), que contienen un número infinito de elementos que varían continuamente de uno a otro, son a la vez variedades y grupos por sí mismos. Estos grupos se denominan **grupos de Lie**, en honor al matemático noruego Sophus Lie. (Qué quiere que le diga, la Noruega del siglo XIX era una especie de hervidero de teoría de grupos). Se pueden contrastar con **grupos discretos** como D_3 o $\mathbb{Z}$, donde los elementos parecen puntos individuales desconectados. Ambos tipos son importantes en física, pero es con los grupos de Lie donde las cosas empiezan a ser realmente profundas.

GRUPOS ORTOGONALES Y ESPACIOS VECTORIALES

Como ya hemos mencionado, el grupo de simetrías del círculo O(2) se llama «grupo ortogonal», ya que también puede considerarse como

«las simetrías de un plano bidimensional, manteniendo el origen fijo y los ejes ortogonales entre sí». Un plano bidimensional es un ejemplo de espacio vectorial. En las aplicaciones reales de la física, pensar en términos de simetrías de espacios vectoriales será mucho más relevante que la idea intuitiva de las simetrías de un círculo. Tendremos conjuntos de campos cuánticos que definen espacios vectoriales de varias dimensionalidades; por ejemplo, los campos de quarks se encuentran en un espacio vectorial «de color» tridimensional con ejes etiquetados como {rojo, verde, azul}. Y muchos de esos valores de campo van a estar descritos por números complejos en lugar de reales. Así que vamos a pensar en los espacios vectoriales de una manera más sistemática.

Ya hemos visto antes los **espacios vectoriales**: conjuntos de elementos, llamados vectores, que pueden sumarse y escalarse mediante números. Todo espacio vectorial tiene un origen, el vector cero. Este actúa como elemento de identidad cuando pensamos en sumar vectores. Por tanto, todo espacio vectorial es en sí mismo un grupo, pero lo que nos interesa aquí son los grupos que transforman los espacios vectoriales, sin pensar en los propios vectores como elementos de grupo.

En *Espacio, tiempo y movimiento*, tratamos generalmente con espacios vectoriales reales (en el sentido de números reales frente a complejos, no en el sentido de que existan físicamente). Si tenemos un espacio vectorial tridimensional con vectores de base $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, un vector general se puede escribir $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3$, donde v_1, v_2 y v_3 son todos números reales, denominados componentes del vector en la dirección adecuada. En otros números de dimensiones, simplemente tenemos más (o menos) vectores de base y componentes: uno por cada dimensión del espacio vectorial. Cuando decimos que los vectores son cosas que pueden «sumarse y escalarse mediante números», en un espacio vectorial real los números mediante los que podemos escalar son números reales. Si $\vec{v}$ y $\vec{w}$ son dos vectores cualesquiera y a y b son números reales, la combinación $a\vec{v} + b\vec{w}$ también es un vector.

Muchos espacios no son vectoriales. Una esfera no es un espacio vectorial, como tampoco lo es el tipo de variedad curva que tenemos en la relatividad general. En ambos casos no hay forma de sumar dos puntos, ni existe un punto preferido llamado «el origen». Los espacios vectoriales son conjuntos continuos de puntos que se extienden infinitamente desde un origen en una geometría plana, generalizaciones de líneas y planos.

Cuando tenemos dos vectores $\vec{v}$ y $\vec{w}$, otro tipo de operación (definida en algunos espacios vectoriales, no necesariamente en todos) es el **producto interno** (también conocido como «producto punto» o «producto escalar»), dado por

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + v_3 w_3. \tag{8.7}$$

Si el ángulo entre los dos vectores es θ, esta fórmula resulta equivalente a

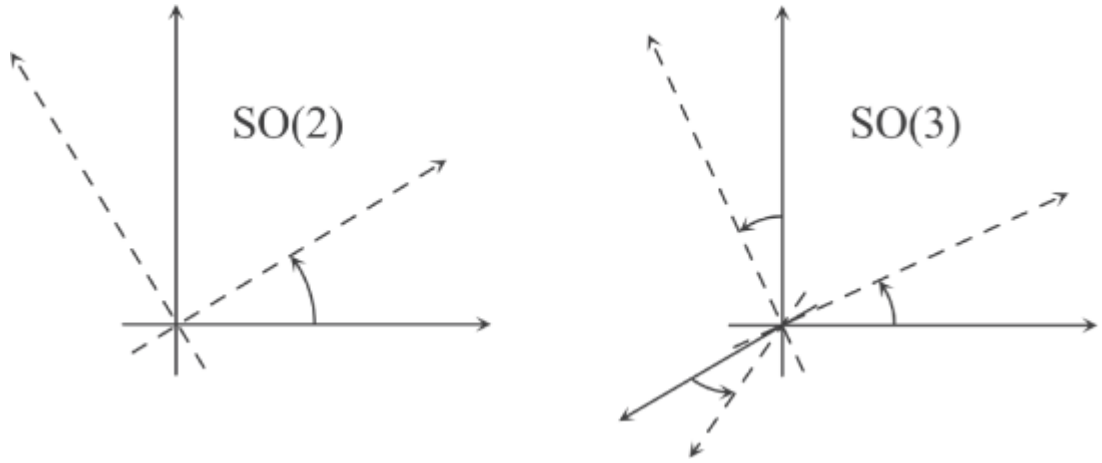
$$\vec{v} \cdot \vec{w} = |v||w| \cos\theta, \tag{8.8}$$

donde $|v|$ y $|w|$ son las longitudes de los vectores respectivos. Cuando dos vectores son ortogonales, el ángulo entre ellos es 90° = $\pi/2$ radianes, y el coseno es 0. Por tanto, el producto interno entre dos vectores ortogonales siempre desaparece.

Una forma de pensar en los grupos ortogonales O(n) es que son transformaciones en un espacio vectorial de n dimensiones que conservan el producto interno entre dos vectores cualesquiera. Esto implica «mantener el origen fijo», porque si movemos el origen, la longitud de los dos vectores cambiará. Por tanto, las transformaciones ortogonales pueden considerarse rotaciones rígidas y reflexiones de vectores n-dimensionales.

El grupo O(3), por ejemplo, es el conjunto de transformaciones que podemos hacer en tres dimensiones que preserva el producto interno (8.6). En otras palabras, que mantiene los ejes ortogonales y el

origen fijo. Equivalentemente, es el grupo de simetrías de la dos-esfera S^2. Este grupo también consta de rotaciones —independientes alrededor de cualquiera de los tres ejes en el espacio tridimensional—, así como de reflexiones. Existe un subgrupo de rotación pura, el grupo ortogonal especial SO(3), y de nuevo un conjunto de grupos $\mathbb{Z}_2$ que invierten la orientación del espacio.



En general, tanto O(n) como SO(n) serán grupos de Lie de dimensión ½ n (n–1), por lo que SO(3) es tridimensional.[2]

El grupo SO(2), de las rotaciones en dos dimensiones, constituye un grupo abeliano. No importa si primero se rota un ángulo θ_1 seguido de un ángulo θ_2, o al revés: el resultado total es una rotación en $\theta_1 + \theta_2$. Pero todos los demás SO(n) para n >2 son no abelianos. En tres dimensiones, *sí* importa si primero se rota alrededor del eje x y luego alrededor del eje y, o viceversa. Por eso los grupos no abelianos son tan importantes en física: a menudo hacemos cosas parecidas a rotaciones entre más de dos ejes.

Paréntesis adicional: hemos estado pensando en elementos de SO(n) en términos de lo que hacen, es decir, rotar cosas en n dimensiones. A menudo, a los físicos les gusta concretar un poco más las cosas y representar tales rotaciones como matrices $n \times n$ que implementarían dicha rotación actuando sobre vectores. Por ejemplo,

un elemento de SO(2) que rota vectores bidimensionales un ángulo q puede escribirse

$$R(\theta)=\begin{pmatrix} \cos\theta & -\text{sen}\theta \\ \text{sen}\theta & \cos\theta \end{pmatrix}. \tag{8.9}$$

En esta forma de pensar, SO(n) es el conjunto de todas las matrices M $n \times n$ que satisfacen la «condición de ortogonalidad», $M^{\mathrm{T}}M = \mathbb{I}$, donde el superíndice T denota la trasposición de la matriz (intercambio de filas con columnas). Del mismo modo, los elementos de los grupos unitarios que consideramos más abajo serán matrices que satisfagan la «condición de unitariedad», que es lo mismo que la condición de ortogonalidad, salvo que también tomamos la compleja conjugada de la primera matriz. Afortunadamente, nunca nos preocuparemos por las formas matriciales explícitas de ninguna de nuestras transformaciones. Para nosotros, los elementos de SO(n) son simplemente «rotaciones en n dimensiones».

REPASO DE LOS NÚMEROS COMPLEJOS

Los números complejos están presentes en toda la física, y la teoría cuántica de campos no es una excepción. El siguiente paso en nuestro viaje por la teoría de grupos es, por lo tanto, actualizar desde los grupos ortogonales SO(n), que representan rotaciones en n dimensiones, a los grupos unitarios SU(n), que representan rotaciones en n dimensiones *complejas*. Lo que significa que necesitamos saber qué se supone que es una «dimensión compleja». Lo que nos da la oportunidad de tomarnos un respiro y recordar algunas cosas sobre los números complejos.

Un número complejo es una combinación de un número real y un número imaginario, como ya hemos visto en los capítulos 1 y 2. Si definimos la unidad imaginaria como $i = \sqrt{-1}$, tenemos un conjunto

completo de números imaginarios que son simplemente números reales habituales multiplicados por i.

Entonces, un número complejo es la suma de un número real y uno imaginario:

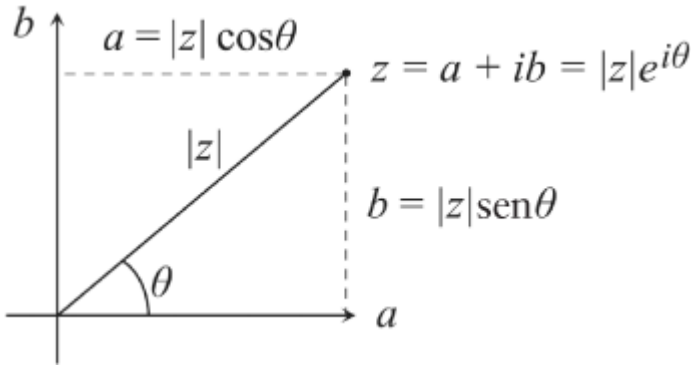
$$z = a + ib, \tag{8.10}$$

donde a y b son a su vez números reales, denominados partes real e imaginaria de z. Cuando sumamos números complejos, las partes real e imaginaria se suman por separado: $(a + ib) + (c + id) = (a + c) + i(b + d)$.

Los números complejos son números: se pueden sumar, multiplicar, dividir, y así sucesivamente. Se puede pensar en ellos y estudiar sus propiedades por derecho propio, sin descomponerlos en una suma de números reales e imaginarios. La principal diferencia es que hay una nueva operación, la **conjugación compleja**, denotada z^*, que es la inversa de sí misma: $(z^*)^* = z$. En términos de las partes real e imaginaria, la conjugación compleja pone un signo menos delante de la parte imaginaria, $z^* = a–ib$. Los números reales permanecen fijos, mientras que $i^* = –i$.

Una buena manera de visualizar los números complejos es en términos del **plano complejo**, como vimos brevemente en el capítulo 1. Se trata simplemente de un plano bidimensional ordinario en el que el eje horizontal representa la parte real del número complejo y el eje vertical representa la parte imaginaria. En esta imagen, la conjugación compleja es solamente una reflexión respecto al eje real, $b \rightarrow -b$. El módulo $|z|$, que no es más que la distancia entre 0 y z en el plano complejo, es un número real que satisface

$$|z|^2 = z^* z = a^2 + b^2. \tag{8.11}$$



En lugar de especificar un número complejo en términos de sus partes real e imaginaria, podemos utilizar alternativamente una «representación polar», en términos del módulo $|z|$ y la **fase** θ del número, como

$$z = |z|\, e^{i\theta}. \tag{8.12}$$

La fase no es más que el ángulo entre el eje horizontal (real) y la línea que va del origen a z en el plano complejo. Podemos obtenerla recordando la fórmula de Euler,

$$e^{i\theta} = \cos\theta + i\,\operatorname{sen}\theta. \tag{8.13}$$

Si observamos la figura del plano complejo y recordamos cómo se utilizan los senos y cosenos para relacionar la longitud de un segmento con sus componentes horizontal y vertical, vemos que (8.13) es justo lo que necesitamos para hacer que la representación polar (8.12) sea correcta.

Por último, podemos pasar a las dimensiones complejas. Un **espacio vectorial complejo** es exactamente lo mismo que un espacio vectorial real, salvo que ahora las componentes son números complejos, y del mismo modo los vectores se pueden escalar mediante números complejos. Si $\vec{\phi}$ es un vector en un espacio vectorial complejo tridimensional, se puede escribir en términos de componentes y vectores unitarios como

$$\vec{\phi} = \phi_1 \vec{e}_1 + \phi_2 \vec{e}_2 + \phi_3 \vec{e}_3, \tag{8.14}$$

donde ϕ_1, ϕ_2 y ϕ_3 son números complejos. (Estas tres dimensiones describen un espacio vectorial complejo abstracto, no el real de tres dimensiones en el que vivimos). Del mismo modo, para cualquier número complejo α y β y los vectores complejos $\vec{\phi}$ y $\vec{\omega}$, la combinación $\alpha\vec{\phi} + \beta\vec{\omega}$ es otro vector complejo. Al igual que en los espacios vectoriales reales, podemos definir un producto interno complejo, pero con un matiz: en lugar de simplemente multiplicar las componentes en cada dirección y luego sumar los resultados, primero tomamos el conjugado complejo de las componentes del primer vector. Así, el producto interno de dos vectores complejos $\vec{\phi}$ y $\vec{\omega}$ se puede escribir

$$\langle \phi, \omega \rangle = \phi_1^* \omega_1 + \phi_2^* \omega_2 + \phi_3^* \omega_3. \tag{8.15}$$

Diferentes referencias pueden utilizar una notación distinta a la de los corchetes de la izquierda. Lo relevante es que tenemos que tener claro para qué vector se está obteniendo su complejo conjugado: el orden es importante.

Por supuesto, ya hemos conocido un importante espacio vectorial complejo: el espacio de Hilbert. Las funciones de onda son vectores complejos. Generalmente usamos una notación especial para ellas, como la notación de «kets» $|\Psi\rangle$ presentada en el capítulo 2, entre otras cosas, para seguir recordándonos las características especiales de los vectores complejos.

Pero la función de onda no es nuestra principal preocupación en este momento. Las simetrías que nos importan van a ser las existentes entre los campos a partir de las cuales construimos nuestra función de onda cuántica. Hemos estado hablando principalmente de un campo escalar real, $\phi(x, t)$. Pero muchos de los campos que conocemos en la física de partículas son complejos y no reales. Ahora estamos en condiciones de hablar de las simetrías de esos campos.

GRUPOS UNITARIOS

El conjunto de los números reales es un espacio vectorial unidimensional. Cada número es un vector; puede sumarse a otros números reales o multiplicarse por algún número real. Pero no hay «rotaciones» de una sola dimensión real. Es solamente una línea, así que ¿qué hay que rotar? Hay un giro: puede enviar cualquier número real x a $-x$. Eso es una única transformación discreta que es su propia inversa. Así que hay un grupo ortogonal en una dimensión, pero es bastante simple: $O(1) \cong \mathbb{Z}_2$.

Una sola dimensión compleja es otra historia. Los números complejos se encuentran en un plano, no en una línea. Y en un plano podemos hacer rotaciones. El **grupo unitario** $U(n)$ es el conjunto de transformaciones de un espacio vectorial complejo n-dimensional que rota este espacio vectorial alrededor del origen. Más formalmente, es el conjunto de transformaciones que dejan invariante el producto interno (8.15). Básicamente son rotaciones rígidas, solamente que en un espacio vectorial complejo. Cuando rotamos todos los ejes al mismo tiempo, el producto interno permanece inalterado.[3]

A diferencia de $O(1)$, que no es más que una reflexión, $U(1)$ es por tanto un conjunto de rotaciones en el plano complejo. De hecho, como grupos, $U(1)$ y $SO(2)$ —el grupo de rotaciones en el plano real bidimensional— son isomorfos: $U(1) \cong SO(2)$. Pero en la práctica parecen diferentes, ya que $U(1)$ actúa sobre números complejos y $SO(2)$ actúa sobre vectores bidimensionales. Ambos son «rotaciones en un plano bidimensional», pero en el caso de $U(1)$ estamos pensando en dos dimensiones reales como una única dimensión compleja. $U(1)$ es un grupo abeliano de Lie unidimensional, lo que sabemos que debe ser cierto porque $SO(2)$ lo es.

No es difícil ser explícito sobre lo que $U(1)$ hace a los números complejos: volviendo a nuestra imagen del plano complejo, un elemento de $U(1)$ simplemente rota la fase del número complejo algún

ángulo w. Eso se consigue multiplicando por $e^{i\omega}$. Si el número con el que empezamos es $z = |z|e^{i\theta}$, el número transformado es $e^{i\omega}z = |z|e^{i(\theta+\omega)}$.

De este modo tenemos los **grupos unitarios especiales** SU(n), que son simplemente rotaciones en n dimensiones complejas. Se trata de grupos de Lie no abelianos de (n^2–1) dimensiones. En forma matricial, un elemento general de SU(2), el grupo unitario especial más simple, puede escribirse

$$M = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \tag{8.16}$$

donde a y b son números complejos que satisfacen $|a|^2 + |b|^2 = 1$. SU(2) se supone que es tridimensional, y de hecho se necesitan tres números reales para especificar dicha matriz: los números complejos a y b son equivalentes a dos números reales cada uno, y hay una restricción en la suma de sus cuadrados, por lo que hay tres parámetros independientes en total. Esta matriz puede utilizarse para rotar vectores en un espacio vectorial de dos dimensiones complejas.

Se podría pensar, siguiendo lo que hicimos con los grupos ortogonales, que primero deberíamos discutir U(n), y luego especializarnos en el caso de rotación pura de SU(n), pero eso no es realmente necesario. Siempre que queramos hacer una transformación U(n), se puede implementar como una transformación SU(n) junto con una transformación U(1) separada. En la práctica, los físicos suelen analizar las simetrías SU(n) y las simetrías U(1) por separado.

Ya está. Ha sido un largo viaje, pero ahora ya conoce los fundamentos de los grupos de simetrías que aparecerán en la física de partículas. Nuestros favoritos serán SO(n), SU(n) y U(1) ≅ SO(2). Son muchas cuestiones matemáticas solamente para sentirse cómodo con la idea de «rotaciones en n dimensiones (reales o complejas)». Pero el esfuerzo merece la pena. Puede llegar bastante lejos como físico profesional solamente con esos grupos.

1 En sentido estricto, hay otros dos axiomas de grupo. En primer lugar, la clausura: la combinación de dos elementos cualesquiera del grupo da otro elemento del grupo. Y en segundo lugar, la asociatividad: para cualesquiera tres elementos del grupo a, b y c, tenemos que $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

2 Elija un eje para moverlo, y otro eje hacia el que rotarlo. Eso es n ejes para empezar, y otros $n-1$ hacia los que girarlo. Así que parecen $n(n-1)$ maneras de rotar, pero elegir los mismos ejes en el orden opuesto no cuenta como una clase separada de rotación, por lo que dividimos entre dos. Siéntase libre de girar alrededor de cualquier línea que desee pasando por el origen, no solamente uno de los ejes de coordenadas; se puede construir una rotación arbitraria como una combinación de rotaciones apropiadas alrededor de los ejes, por lo que estos son suficientes para definir la dimensionalidad del grupo.

3 Punto técnico: se podría pensar que la conjugación compleja $z \to z^*$ es análoga a un giro de las dimensiones complejas y, por tanto, debería contar como una transformación unitaria. No es así; para verlo, piense en el producto interno (8.15). Este es un número complejo, y en general también se conjugará cuando conjuguemos ambos vectores. Pero una transformación unitaria se define para dejar el producto interno completamente inalterado, por lo que la conjugación compleja no cuenta.

NUEVE

TEORÍA DE CAMPO DE GAUGE

En lo que llevamos de libro, hemos acumulado un impresionante repertorio de conocimientos técnicos sobre mecánica cuántica y teoría cuántica de campos. Hemos aprendido sobre funciones de onda, por qué los campos dan lugar a partículas al cuantizarse, cómo esas partículas interaccionan entre sí y cómo hablar de simetrías. En este capítulo y en los dos siguientes, nuestra acumulación de conocimientos alcanza un *crescendo* cuando nos enfrentamos a las teorías de campo de gauge y a la distinción entre fermiones y bosones, dos pilares de la física fundamental moderna. Todo ello nos conducirá a una muy merecida recompensa en el último capítulo, en el que reuniremos los ingredientes del Modelo Estándar de la física de partículas y utilizaremos algunas ideas sencillas (leyes de conservación, partículas pesadas que se desintegran en otras ligeras) para explicar características fundamentales del mundo que vemos a nuestro alrededor.

Este capítulo y el siguiente tratan sobre las **teorías de campo de gauge**, que son un tipo particular de teoría de campos con un tipo particular de simetría: una transformación que puede realizarse independientemente en cada punto del espaciotiempo. Esta simple idea va a tener enormes consecuencias. Vamos a tener que introducir campos adicionales para que esta simetría expansiva funcione, y esos campos van a acabar siendo los campos portadores de fuerzas de la física de partículas: fotones, gluones, bosones W y Z y gravitones. La simetría de gauge es un poderoso principio en el que se basan las fuerzas de la naturaleza.

Cuando hablamos por primera vez de cómo las interacciones se describen mediante vértices del diagrama de Feynman que pueden trazarse hasta los términos que aparecen en un lagrangiano de interacción, el proceso puede haber parecido un poco flojo. ¿Quién decide qué campos hay o cómo interaccionan? Con las teorías de campo de gauge hay mucha menos arbitrariedad. Una vez que se tiene un campo y se quiere que sea invariante bajo una simetría de gauge, eso exige la existencia de otros campos, y la forma en que esos campos interaccionan va a estar determinada en gran medida por la simetría. Es una base muy sólida sobre la que construir una teoría física.

En estos dos capítulos habrá más notación matemática que en la mayor parte del resto del libro. Hay una razón para ello: ahora conoce suficientes principios subyacentes como para ver realmente por qué la invariancia gauge implica propiedades tan cruciales como la ausencia de masa del fotón, la existencia de antipartículas y la conservación de la carga. Sería una pena llegar tan lejos y contentarnos con algunos gestos informales. Es más divertido ponerse la gorra de pensar y llegar a una comprensión clara de algunos de los secretos más profundos de la naturaleza.

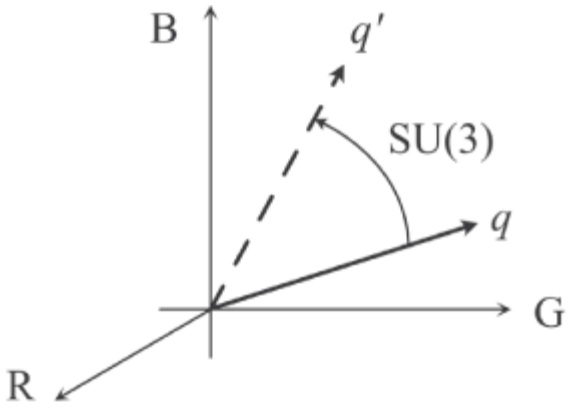
QUARKS Y COLOR

Para ilustrar rápidamente lo que queremos decir, pensemos en un ejemplo del mundo real. Los **quarks**, como vimos en el capítulo 7, son las partículas que existen dentro de los protones, neutrones y otras partículas que interaccionan fuertemente. La fuerza que las mantiene unidas se describe mediante la cromodinámica cuántica (CDC), donde el papel de la «carga» lo desempeña una nueva magnitud llamada «color».

A menudo se dice que los quarks tienen uno de estos tres colores: rojo, verde o azul. Esto no es del todo exacto. Lo que ocurre es que cada campo q de los quarks está en realidad, en cada punto del

espacio-tiempo, descrito por un vector que existe en un espacio vectorial complejo tridimensional —el **espacio de color**— y son los tres ejes de ese espacio los que se denominan rojo (R), verde (G) y azul (B). El espacio de color es un **espacio vectorial interno**, no tiene nada que ver con la velocidad o el momento ni con ningún otro vector familiar del espacio ordinario o el espacio-tiempo. Es un espacio que simplemente existe en cada punto, y cada campo de quark (el quark *up*, el quark *down*, y así sucesivamente) define un vector particular dentro de ese espacio. Dicho de otro modo, hay tres colores básicos de campos de quark (R, G, B), y algo como «el campo del quark *up* en (*x*, *t*)» será una combinación vectorial de los tres, con componentes diferentes en cada dirección. Los vectores de color tienen una dirección en el espacio de color, no en el espacio real.

No le sorprenderá saber que existe una simetría SU(3) actuando en este espacio complejo tridimensional. Podemos «rotar» el campo de quark con una transformación SU(3), que cambia las cantidades relativas de las componentes R, G y B del campo mientras deja su longitud inalterada. Citamos «rotar» entre comillas porque no se trata de una rotación literal en el espacio, sino solamente una transformación entre tres direcciones en este espacio de color interno.



Pero a la naturaleza no le importa cómo expresemos el vector del quark en términos de componentes, o cómo estén orientados los ejes: esto es solamente una elección que hacemos por conveniencia, no una

característica fundamental del universo. Por eso hay una simetría. La forma en que orientamos el vector campo con respecto a los ejes es físicamente irrelevante, y las simetrías a menudo ponen de manifiesto el hecho de que existen múltiples formas equivalentes de representar la misma situación física. Para un campo de un único quark, lo único que importa es la longitud del vector; cuando empecemos a comparar diferentes campos, o un único campo en diferentes puntos del espacio-tiempo, el ángulo relativo entre los vectores de color empezará a importar. Estas propiedades —longitud de vectores individuales, ángulos entre vectores— son precisamente lo que una transformación SU(3) deja invariante.

TRANSFORMACIONES DE GAUGE

Pero eso no es lo más interesante. Lo importante de la simetría de color SU(3) es que no es una **simetría global**, sino una **simetría de gauge**, también conocida como **simetría local**. Aquí es donde ocurre la magia.

Las transformaciones de simetría global son aquellas que transforman el campo exactamente de la misma manera en todos los puntos del espacio-tiempo. Por este motivo, a veces se denominan simetrías «rígidas». Denotemos a rotación de color particular por un elemento de grupo SU(3) específico G. Una transformación de los campos de quarks puede escribirse

$$G\colon q\,(x,\,t)\rightarrow q'\,(x,\,t) = G\cdot q\,(x,\,t). \qquad (9.1)$$

La notación del lado derecho solamente significa «aplicar la transformación G sobre el campo q».

En cambio, una simetría de gauge implica una transformación que ocurre independientemente en cada punto. En lugar de un elemento de grupo fijo G, consideramos una transformación dependiente del

espacio-tiempo $G(x, t)$. La regla sigue pareciéndose a (9.1), pero la transformación de q a q' puede ser diferente en cada punto.

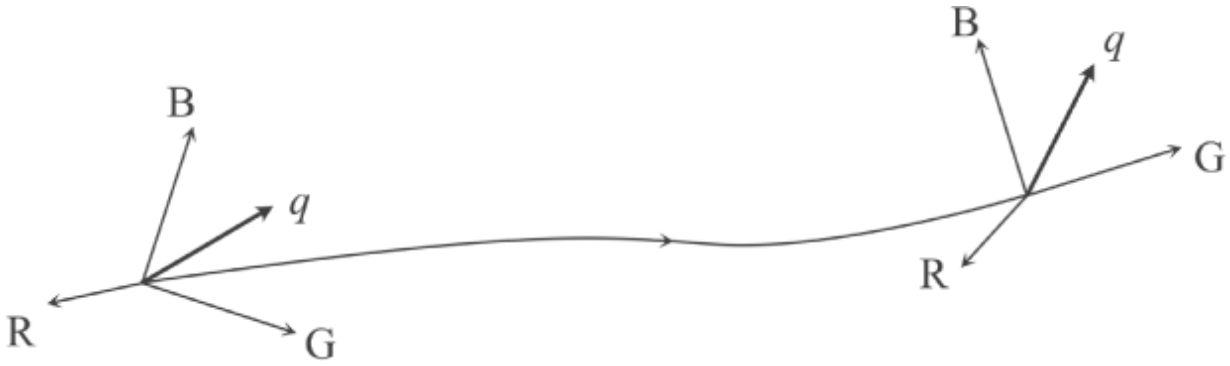
$$G(x, t): q\ (x, t) \rightarrow q'\ (x, t) = G\ (x, t) \cdot q\ (x, t). \quad (9.2)$$

Se trata de una **transformación de gauge**, una transformación de simetría que depende del lugar del espacio y el tiempo en el que nos encontremos.

Puede que esto no parezca gran cosa. Dada la idea básica de que «la dirección de los ejes RGB es físicamente irrelevante», es natural que sea irrelevante por separado en cada punto del espacio-tiempo.

Y lo es. Pero surge un nuevo problema. Mientras que la dirección en el espacio de color a la que apunta el campo en un punto concreto es una cuestión de convención, la dirección *relativa* a la que apunta el campo en dos puntos distintos del espacio sí que puede ser bastante importante. Queremos ser capaces de decir si el campo de quarks en dos puntos diferentes está haciendo lo mismo o algo distinto. (Por ejemplo, queremos saber que todas las excitaciones de los quarks dentro de un protón tienen un color total de cero).

Pero ¿cómo vamos a comparar lo que hace el campo en dos puntos diferentes del espacio si tenemos libertad para girar nuestros ejes por separado en cada punto? Dadas dos localizaciones cualesquiera, siempre podemos hacer una transformación de gauge como en (9.2) para hacer que el campo de quarks apunte a lo largo de «azul» en ambos lugares o, para el caso, hacer que el campo sea rojo en un punto y verde en el otro. Necesitamos alguna forma de comparar realmente los campos en diferentes puntos, una forma que no se vea afectada por las transformaciones de gauge que decidamos hacer.



CONEXIONES

Ya nos enfrentamos a un problema similar en el espaciotiempo curvo de la relatividad general, pero allí pensábamos en los vectores de toda la vida que apuntan en una dirección en el espacio-tiempo, en lugar de los vectores internos del espacio de color. En el espacio-tiempo plano, podríamos tener un vector en un lugar y otro vector en otro lugar, y simplemente compararlos entre sí, al menos, mientras se opte sensatamente por utilizar un sistema de coordenadas ortogonales (cartesianas) en todas partes. Pero una vez que el espacio-tiempo se curva, no existe nada parecido a un sistema de coordenadas ortogonal en todas partes. Tenemos que pensar más detenidamente en cómo transportar un vector de un punto a otro, a lo largo de una trayectoria determinada, para poder llevar dos vectores a un punto común y compararlos.

Los lectores atentos de *Espacio, tiempo y movimiento* tal vez recuerden que, en efecto, podemos hacerlo, y la respuesta va a implicar otro tipo de campo. Existe una **ecuación de transporte paralelo** que nos permite mover vectores (u otros tensores) por cualquier trayectoria que queramos, manteniéndolos tan constantes como podamos a lo largo del camino. Pero necesitamos información matemática que nos diga cómo cambiarían adecuadamente las componentes del vector durante el proceso. Esa información viene en forma de **conexión**, que es en sí misma un tipo de campo. En

relatividad general, resulta que la conexión puede derivarse del campo tensorial métrico, razón por la que no tuvimos que especificarla por separado. Es posible pensar en la relatividad general como una teoría de campo de gauge, tal y como describió en primer lugar Ryoyu Utiyama en 1956, siendo el grupo de gauge el grupo de Lorentz SO(3,1). («Rotaciones» en un espacio-tiempo con tres dimensiones espaciales y una temporal).

He aquí una analogía posiblemente útil: imagine que está construyendo un robot que trabajará en un restaurante llevando una bandeja cargada de vasos llenos de bebidas. Quiere asegurarse de que los vasos siempre estén orientados exactamente en vertical para que el robot no derrame ni una sola gota. Pero el restaurante está al aire libre y en un terreno accidentado, por lo que la bandeja necesita ajustarse constantemente para permanecer horizontal. Y (no me mire así, yo simplemente me invento las analogías) el robot no tendrá sensores ni otros dispositivos que le permitan evaluar la altura y el ángulo del terreno que pisa. En lugar de ello, le va a dar esa información por adelantado: la topografía exacta de la zona por la que podría desplazarse el robot, para que sepa cómo ajustar continuamente su bandeja para evitar que se derramen las bebidas. Esa información que le está dando al robot es esencialmente una conexión: es un campo (información sobre algo que ocurre en cada punto del espacio) que permite al robot poner en práctica la idea de «mantener la bandeja exactamente horizontal y, por tanto, los vasos de bebida exactamente verticales» mientras se desplaza. El campo de conexión es precisamente el conjunto de datos que «conectan» un punto con lo que ocurre en otros puntos.

Algo similar ocurrirá en nuestras teorías de campo de gauge con espacios vectoriales internos. Para comparar los valores de los campos de quarks en diferentes lugares, necesitamos introducir un **campo de gauge**, también llamado **potencial vectorial**, que nos permitirá mover los campos de un lado a otro de una forma invariante de gauge. «Potencial vectorial» viene del hecho de que es un campo vectorial,

generalmente escrito $A_\mu(x,t)$. Así que en lugar de un único campo de quarks, en esta teoría de campo de gauge necesitamos dos campos:

$$q(x,t),\ A_\mu(x,t).$$

El tira y afloja entre estos dos tipos diferentes de campos —campos de materia, como quarks y leptones, y campos de gauge, como gluones y fotones— crea nucleones, átomos, moléculas... hasta llegar a usted y a mí.

Matemáticamente hablando, el campo de gauge es una conexión, muy parecida a la que utilizábamos para el transporte paralelo en relatividad general. Proporciona la información que necesita para decir, por ejemplo, «el campo de quarks es constante a lo largo de esta trayectoria», incluso aunque las transformaciones de gauge le permitan cambiar la orientación de sus ejes rojo/verde/azul como quiera de un punto a otro.

Hay complicaciones cruciales y divertidas cuando tenemos simetrías de gauge no abelianas como SU(3), en lugar de una abeliana como U(1). En ese caso, el campo de gauge A_μ no es simplemente un vector, sino un vector matricial con índices adicionales que le permiten realizar rotaciones en el espacio de color o cualquier otro espacio vectorial interno que estemos considerando. Para simplificar, no vamos a escribirlos y, en su lugar, utilizaremos A_μ para cualquier campo de gauge del que estemos hablando en ese momento. Las ideas fundamentales siguen siendo las mismas.

Si todo esto le parece un poco confuso, recuerde lo esencial: para tener invariancia de gauge, necesitamos un nuevo campo, el campo de gauge. Si no quiere preocuparse por el transporte paralelo, las conexiones y los espacios vectoriales internos, limítese a la necesidad de asegurarse de que todo sea realmente invariante bajo la simetría de gauge.

Una vez que hemos reconocido la necesidad de un campo de gauge o de conexión, hay muchas razones para pensar en ese nuevo campo

por derecho propio. ¿Cuál es su dinámica? ¿A qué tipo de partículas da lugar cuando lo cuantizamos? ¿Cómo interaccionan esas partículas con las de otros campos?

Cada simetría de gauge viene acompañada de su propio campo de conexión, y esos campos y sus interacciones son lo que dan lugar a las fuerzas conocidas de la naturaleza. Las partículas asociadas a esos campos se denominan **bosones de gauge**. Para SU(3) del color, los bosones de gauge son gluones. Para electromagnetismo, se trata de fotones. Para las interacciones débiles se trata de los bosones W y Z, pero, como veremos, hay nuevas complicaciones debido al bosón de Higgs y a su truculencia. Nadie dijo nunca que la física de partículas fuera algo claro y ordenado.

Las simetrías no son solamente una simplificación cómoda o una atractiva característica estética de los campos cuánticos. Aplicarlas de forma coherente conduce directamente a fuerzas entre partículas de materia. Esta idea es la base del Modelo Estándar de la física de partículas.

INVARIANCIA DE GAUGE

Será útil dar un paso atrás en las complicaciones de la CDC y examinar un caso más sencillo: la EDC, la electrodinámica cuántica, la teoría de los electrones, positrones y fotones. Como en la CDC, existe un campo de gauge A_μ que nos ayuda a comparar campos en diferentes puntos. Este hecho tiene implicaciones cruciales para las propiedades de este campo y cómo interacciona.

Respiremos hondo y pensemos cómo se supone que se comportará este campo cuando hagamos una transformación de gauge. Dicho de forma abreviada: el trabajo de A_μ es asegurar que todo permanezca invariante de gauge, y ese requisito determinará cómo se transforma el propio A_μ. Y esas propiedades de transformación implicarán que las características físicamente relevantes («invariantes de gauge») son en

realidad derivadas de A_μ, y esas derivadas resultan ser los conocidos campos eléctricos y magnéticos (o su generalización a otras fuerzas).

Tomemos un campo cargado eléctricamente como el del electrón. Lo denotaremos por $\psi_e(x,t)$, ya que la letra e por sí sola ya está tomada para la constante de Euler, que usaremos mucho. (Pero ψ_e aquí es un campo, no una función de onda). Es un campo de valores complejos, no reales. Dado que los electrones son partículas de espín ½, no se trata de un campo escalar, sino de algo llamado **espinor**, lo que lleva a todo tipo de complicaciones, que ignoraremos alegremente.

Lo que nos importa es que ψ_e es un campo complejo. Eso le permite tener una simetría de gauge U(1) y, de hecho, la tiene.

$$\mathrm{U}(1):\ \psi_e(x,t) \rightarrow \psi'_e(x,t) = e^{i\theta(x,t)}\psi_e(x,t). \qquad (9.3)$$

Todo lo que hacemos es multiplicar el campo complejo por un factor de fase dependiente del espacio-tiempo, rotando el valor del campo en el plano complejo un ángulo $\theta(x, t)$. Queremos insistir en que la teoría en su conjunto sea invariante bajo cualquier tipo de transformación de simetría. Por lo tanto, vamos a necesitar la introducción de un campo de gauge $A_\mu(x,t)$. Este actuará como un «potencial» para los campos eléctrico y magnético, y su cuantización dará lugar a fotones, por lo que también podemos llamarlo el **campo del fotón**.

El campo de gauge A_μ es un campo de valores reales, no complejos. Sin embargo, para que cumpla su función (decirnos cómo realizar un transporte paralelo del campo del electrón de un lugar a otro), también tendrá que cambiar cuando hagamos una transformación de gauge. Si no fuera por la simetría de gauge, la simple idea de «el campo del electrón es constante en todas partes» se implementaría estableciendo que sus derivadas parciales con respecto a las coordenadas sean iguales a 0 (usando la notación $\partial/\partial x^\mu = \partial_\mu$):

$$\partial_\mu \psi_e = 0. \tag{9.4}$$

Bastante sencillo, pero hay un problema inmediato cuando hacemos una transformación de gauge. Si dejamos que $\psi_e \to e^{i\theta}\psi_e$ y luego tomamos la derivada, obtenemos (manteniendo la derivada de la propia ψ_e en 0)[1]

$$\partial_\mu \psi_e \to \partial_\mu\left(e^{i\theta}\psi_e\right) = \left(\partial_\mu e^{i\theta}\right)\psi_e + e^{i\theta}\left(\partial_\mu \psi_e\right) = i\left(\partial_\mu \theta\right)e^{i\theta}\psi_e. \tag{9.5}$$

Vemos que la derivada transformada de gauge del campo del electrón no sigue siendo 0, sino que es proporcional a la derivada de θ. Así que desaparecería si la propia θ fuera constante a través del espacio-tiempo ($\partial_\mu\theta = 0$), pero en ese caso tendríamos simplemente una transformación global rígida en lugar de una transformación de gauge dependiente del espacio-tiempo. Únicamente tenemos una verdadera simetría si la cantidad permanece completamente inalterada bajo la transformación. ¿Cómo podemos arreglar las cosas de modo que sigamos teniendo una simetría cuando θ cambie de un lugar a otro?

En una teoría de campo de gauge, la simple ecuación (9.4) no es lo que queremos decir con «mantener el campo constante». Como dijimos, necesitamos utilizar el campo de gauge para dar sentido a esa afirmación, ya que necesitamos comparar valores de ψ_e en distintos lugares. La forma de hacerlo es sustituir (9.4) por

$$\partial_\mu \psi_e - iA_\mu \psi_e = 0. \tag{9.6}$$

Esta no es una ecuación que se suponga que siempre es cierta; es solamente la generalización de la teoría de campo de gauge de la condición $\partial_\mu\psi_e = 0$. En lugar de fijar en 0 las derivadas parciales del

campo, hay que compensarlas con el campo multiplicado por el potencial vectorial. La expresión del lado izquierdo va a desempeñar un papel crucial en lo que viene a continuación. A veces se la llama **derivada covariante de gauge** de ψ_e, pero esa rimbombante nomenclatura es menos importante que la idea subyacente de corregir una derivada mediante la conexión. (Piense en mantener el campo constante como «sujetar la bandeja horizontalmente», y el segundo término como «cómo compensar el hecho de que el suelo está inclinado»).

La ventaja de (9.6) respecto a (9.4) es que no se ve modificada por una transformación de gauge. Para que ocurra, necesitamos un tipo específico de regla de transformación para el propio A_μ:

$$\mathrm{U}(1):\ A_\mu(x,t) \to A'_\mu(x,t) = A_\mu(x,t) + \partial_\mu \theta. \qquad (9.7)$$

Así, mientras que el campo del electrón se multiplica por $e^{i\theta}$, el campo del fotón se desplaza en $\partial_\mu\theta$, la derivada parcial de θ con respecto a las coordenadas del espacio-tiempo. Esto va a tener consecuencias cruciales más adelante. Y tiene sentido. La misión de A_μ es compensar una transformación de gauge dependiente del espacio-tiempo. Pero si θ es simplemente una constante, se trata de una rotación global que no cambia de un lugar a otro. En tal caso, la derivada desaparece y A_μ no varía. Así que cualquier cambio que experimente va a depender de $\partial_\mu\theta$, y el cambio correcto resulta ser simplemente aditivo.

Ahora volvamos a (9.6). Cuando hacemos una transformación de gauge U(1), el campo del electrón se multiplica por $e^{i\theta}$ como en (9.3), por lo que su derivada cambia como en (9.5). Pero el campo del fotón también se transforma, desplazándose en $\partial_\mu\theta$ según (9.7). El resultado es que (9.6) solamente se multiplica por una fase global $e^{i\theta}$ después de una transformación de gauge. Por lo tanto, si esta cantidad es 0 al principio, seguirá siendo 0.

Cuando James Clerk Maxwell escribió las ecuaciones del electromagnetismo clásico en el siglo XIX (subido a hombros de gigantes, por cierto), lo hizo en términos del campo eléctrico $\vec{E}$ y del campo magnético $\vec{B}$, ambos vectores tridimensionales ordinarios. ¿Cómo se relacionan con el campo de gauge del que estamos hablando? La respuesta es que son sus derivadas, pero de una forma muy específica. Si me permiten volver a incidir brevemente en el lenguaje tensorial del que hablamos en *Espacio, tiempo y movimiento*, el **tensor de intensidad de campo** $F_{\mu\nu}$ es un tensor de dos índices, dado por

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu . \qquad (9.8)$$

Nótese que los índices μ y ν se han intercambiado entre los dos términos. Si no, obtendríamos simplemente 0. A diferencia del tensor métrico, que es simétrico ($g_{\mu\nu} = g_{n\mu}$), el tensor de intensidad de campo es **antisimétrico**: $F_{\mu\nu} = - F_{\mu\nu}$.

Y lo que es más importante, el tensor de intensidad de campo es **invariante de gauge**: cuando hacemos una transformación de gauge, $F_{\mu\nu}$ no cambia en absoluto. Podemos ver esto si tomamos la transformación de gauge de A_μ de (9.7) y la introducimos en la definición de la intensidad de campo (9.8), obteniendo

$$\partial_\mu \partial_\nu \theta - \partial_\nu \partial_\mu \theta = 0. \qquad (9.9)$$

Esto es automáticamente cero porque las derivadas parciales siempre conmutan entre sí: obtenemos la misma respuesta independientemente del orden en que las tomemos. Cuando trabajamos con teorías de campo de gauge, gran parte del juego consiste en asegurarse de que las cantidades físicamente relevantes sean invariantes de gauge. El tensor de intensidad de campo $F_{\mu\nu}$ es un ejemplo paradigmático: (9.9) muestra que no cambia cuando

aplicamos (9.7) a A_μ. El propio campo de gauge no es invariante de gauge, pero su correspondiente tensor de intensidad de campo sí lo es.

La relación de $F_{\mu\nu}$ con los campos eléctrico y magnético $\vec{E}$ y $\vec{B}$ es que son la misma cosa. Más concretamente, si pensamos en $F_{\mu\nu}$ como una matriz 4 × 4, los campos eléctrico y magnético son sus componentes:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{9.10}$$

Así, $F_{01} = E_x$, $F_{23} = B_x$, y así sucesivamente. Por eso A_μ se llama a veces el vector «potencial»: al igual que la fuerza gravitatoria en la mecánica newtoniana puede considerarse como la derivada de un campo de potencial gravitatorio, los campos eléctrico y magnético (que dan lugar directamente a fuerzas) pueden considerarse derivadas apropiadas de A_μ. El tensor de intensidad de campo es directamente análogo al tensor de Riemann de la relatividad general: ambos son medidas invariantes de la «curvatura» de los potenciales subyacentes.

A veces se oye decir que el propio Maxwell no conocía la invariancia de gauge, porque trabajaba directamente con los campos eléctrico y magnético. Eso no es cierto en absoluto: Maxwell sabía que se podía definir un potencial vectorial y derivar de él los campos eléctrico y magnético, y que al potencial se le podría aplicar una transformación de gauge (aunque no lo llamó así) sin alterar $\vec{E}$ o $\vec{B}$. No sabía nada de relatividad especial, por supuesto, así que nunca organizó los campos electromagnéticos en un tensor espacio-temporal. Tampoco conocía el campo del electrón: ni siquiera conocía los electrones. Es en el contexto de la teoría cuántica de campos donde el

campo de gauge A_μ y sus propiedades se vuelven realmente indispensables.

EL ELECTRÓN Y EL POSITRÓN

Juntemos todo esto para hacer algo de física.

Si recuerda el capítulo 5, Interacciones, sabrá que por «hacer física» entendemos «escribir un lagrangiano, y a partir de él construir diagramas de Feynman para representar interacciones». La invariancia de gauge va a ser crucial en este proceso. El juego al que vamos a jugar es intentar construir algo similar a nuestro lagrangiano anterior para un campo escalar, pero ahora con el requisito adicional de que la teoría sea invariante bajo simetría de gauge.

La acción es la integral sobre todo el espacio-tiempo de la densidad de Lagrange, $S = \int \mathcal{L}\, dt\, d^3x$, y la densidad de Lagrange —normalmente llamada simplemente «el lagrangiano»— se construye a partir de los campos y sus derivadas. El lagrangiano para campos será una suma de diferentes términos. Hay términos cinéticos y de gradiente con derivadas de los campos con respecto al espacio y al tiempo, normalmente unificados en un «término cinético», ya que el espacio y el tiempo están unificados en la relatividad. También podría haber términos de masa de la forma (campo)^2, y términos de interacción en los que intervienen más de dos potencias de los campos.

$$\mathcal{L} = (\text{término cinético}) + (\text{término de masa}) + (\text{interacciones}). \tag{9.11}$$

Vamos a querer asegurarnos de que el lagrangiano en su conjunto es invariante de gauge. Los términos individuales no necesitan serlo, si se transforman de una manera que se compense con el cambio en algún otro término.

Pensemos primero en el campo del electrón. Tiene un término cinético, escrito por primera vez por Paul Dirac, pero no nos preocupemos de él por el momento. Hay una serie de complicaciones matemáticas derivadas del hecho de que el electrón tiene espín ½, que llevaron a Dirac a plantear sus famosas «matrices de Dirac», que en gran medida vamos a obviar.

Considere el término de masa. Podría suponer que si el campo del electrón es ψ_e, su término de masa debería ser simplemente $(\psi_e)^2$. Pero eso no puede ser correcto. Mire lo que ocurre bajo una transformación de gauge como (9.3):

$$\left(\psi_e\right)^2 \rightarrow \left(e^{i\theta}\psi_e\right)^2 = e^{2i\theta}\psi_e^2. \tag{9.12}$$

Eso no es invariante: agrega un factor de fase $e^{2i\theta}$. ¿Significa esto que los electrones no pueden tener masa después de todo?

Claramente no, ya que los electrones sí tienen masa (0,511 MeV). Simplemente tenemos que pensar con más cuidado cómo hacer un término de masa invariante de gauge. Afortunadamente, nos guardamos un as bajo la manga: como el campo del electrón es complejo, también podemos trabajar con el campo complejo conjugado ψ_e^*. Sabemos lo que le ocurre bajo una transformación de gauge, ya que sabemos lo que le ocurre al propio ψ_e y la conjugación compleja cambia cada i por $-i$:

$$\psi_e^* \rightarrow \left(e^{i\theta}\psi_e\right)^* = e^{-i\theta}\psi_e^*. \tag{9.13}$$

Esto nos permite hacer algo invariante de gauge, a saber, ψ_e * ψ_e. De hecho, sabemos de inmediato que esto es invariante de gauge, ya que es simplemente el módulo al cuadrado $|\psi_e|^2$, y la particularidad más importante de una rotación U(1) es que deja el módulo invariante, mientras desplaza la fase del campo complejo. Pero para estar seguros, también podemos resolverlo:

$$\psi_e^* \psi_e \rightarrow \left(e^{-i\theta}\psi_e^*\right)\left(e^{i\theta}\psi_e\right) = e^{-i\theta+i\theta}\psi_e^*\psi_e = \psi_e^*\psi_e . \tag{9.14}$$

Una transformación de gauge U(1) deja esta combinación completamente inalterada, tal y como esperábamos. Eso es lo que buscamos en un lagrangiano que se comporte bien.

¿Estamos realmente seguros de que está bien utilizar el complejo conjugado del campo original para construir nuestro término de masa? La respuesta es sí, pero los físicos tardaron un tiempo en entender por qué. El truco es pensar en ψ_e^* como un campo en sí mismo, adicional a ψ_e, en lugar de derivarse de él. Y si todo eso es cierto, habrá cuantos de ψ_e^* que veremos como partículas.

Efectivamente, y esa partícula es el **positrón**, la antipartícula del electrón. El razonamiento por el que hemos pasado es una versión simplificada, de juguete, de cómo Dirac dedujo la ecuación que lleva su nombre para el electrón allá por 1928. Su ecuación parecía implicar que debía existir otra partícula con la misma masa que el electrón, pero con carga eléctrica opuesta. Puede que haya oído que esta predicción tenía algo que ver con la comprensión del espín del electrón, e históricamente ha desempeñado un papel importante. Pero, de hecho, cualquier partícula cargada, es decir, con una simetría de gauge U(1), va a tener una antipartícula, independientemente del valor que pueda tener su espín.

El término de masa en el lagrangiano que fija la masa del electrón también hace lo mismo con el positrón, por lo que no tienen más remedio que tener la misma masa. Los físicos se mostraron reacios a aceptar la predicción de una nueva partícula similar al electrón con carga positiva —¿por qué no la habían descubierto ya?—, pero el positrón fue hallado finalmente por Carl Anderson en 1932.

Es posible que haya oído que el campo del bosón de Higgs es el responsable de que las partículas tengan masa. Si es así, ¿por qué hemos podido escribir un término de masa para el electrón sin hablar nunca del bosón de Higgs? La respuesta es que la necesidad del bosón

de Higgs en el Modelo Estándar de la física de partículas es bastante especial. Tiene que ver con el hecho de que las interacciones débiles violan la paridad: tratan a las partículas con espín hacia la izquierda de forma diferente que a las partículas con espín hacia la derecha. Como Dirac no estaba pensando en la paridad ni en las interacciones débiles, su teoría no tuvo ningún problema en incluir el término de masa para el electrón y seguir adelante con su vida.

Mientras tanto, la simetría de la invariancia de gauge es suficiente para prohibir toda masa para otra famosa partícula: el fotón.

FUERZAS DE LARGO ALCANCE

A continuación, vamos a pensar en el lagrangiano para el campo del fotón A_μ. Una vez más, la invariancia de gauge va a desempeñar un papel importantísimo. Para el término cinético, ya conocemos una manera fácil de hacer un tensor invariante de gauge: la intensidad de campo $F_{\mu\nu}$, definida en (9.8). Pero se supone que el lagrangiano es un escalar (cero índices), no un tensor con dos índices. Recordemos que el tensor de intensidad de campo es análogo al tensor de curvatura en relatividad general. En ese caso pudimos construir un escalar R que podía utilizarse para hacer un lagrangiano, pero eso no funciona para el electromagnetismo. Lo que podemos hacer es multiplicar la intensidad de campo por sí misma, para obtener $(F_{\mu\nu})^2$. Y, en efecto, ese es el término cinético correcto para electromagnetismo. La invariancia de gauge nos facilita la vida.

Nota rápida de procedimiento: siempre que multiplicamos tensores entre sí, tenemos que «absorber los índices» sumando sus posibles valores para obtener una cantidad escalar. Esto puede hacerse contrayendo con factores apropiados de la métrica $g_{\mu\nu}$ y la métrica inversa $g^{\mu\nu}$, usándolos para subir y bajar índices y asegurándonos de sumar sobre índices repetidos cuando uno está arriba y otro abajo. En este libro ignoraremos por completo esa sutileza. Escribiremos $(A_\mu)^2$

cuando en realidad queramos decir $A_\mu\, A^\mu$, y del mismo modo $(F_{\mu\nu})^2$ para $F_{\mu\nu}\, F^{\mu\nu}$, y así sucesivamente. Siempre que vea algo así a partir de ahora, asuma que estamos subiendo y bajando índices implícitamente y sumándolos de forma apropiada.

¿Y qué pasa con el término de masa para el fotón? Lo más sencillo sería imaginar algo como $(A_\mu)^2$. Pero eso, evidentemente, no es invariante de gauge, como podemos comprobar introduciéndolo en (9.7):

$$\mathrm{U}(1):\ \left(A_\mu\right)^2 \rightarrow \left(A_\mu + \partial_\mu \theta\right)^2 = \left(A_\mu\right)^2 + 2A_\mu(\partial_\mu\theta) + \left(\partial_\mu\theta\right)^2. \tag{9.15}$$

Este término cambia claramente cuando hacemos una transformación de gauge (al menos, siempre que $\partial_\mu\theta$ no sea cero), así que no es invariante de gauge. Para el electrón podríamos solucionar este problema invocando el campo complejo conjugado. Pero A_μ es un campo vectorial de valores reales, no tiene un complejo conjugado separado. ¿Qué tenemos que hacer?

La respuesta es bien sencilla: renunciar. Sin introducir campos completamente nuevos (análogos al de Higgs) que cambiarían drásticamente la física del electromagnetismo, no hay manera de dar masa al fotón en una forma invariante de gauge. Así pues, vemos otra implicación física crucial de la invariancia de gauge: las partículas portadoras de fuerza carecen naturalmente de masa.

Un razonamiento similar se aplica a la gravedad descrita por la relatividad general. En ambos casos obtenemos fuerzas de largo alcance portadas por partículas sin masa porque hay simetrías que prohíben a las partículas obtener masa, lo que da lugar a leyes de fuerza que varían con la inversa del cuadrado (ley de Coulomb para electromagnetismo, ley de gravitación universal de Newton para la gravedad). Este es el motivo por el que son esas dos fuerzas las que resultan más evidentes en nuestras vidas macroscópicas a escala humana. La energía de una partícula está relacionada con su momento

y su masa por $E^2 = p^2 + m^2$, por lo que la energía mínima que puede tener una partícula (es decir, cuando $p = 0$) es $E = m$. Se necesita una cierta energía mínima para crear una partícula masiva; en términos de las fuerzas que transmiten las partículas, eso implica que esas fuerzas disminuyen rápidamente con la distancia. Por su parte, las partículas sin masa pueden ser tan poco energéticas como queramos. Los gravitones o los fotones virtuales de baja energía no necesitan mucho esfuerzo para atravesar enormes distancias espaciales, lo que hace posibles las fuerzas de largo alcance.

INTERACCIONES DE LA EDC

Así que tenemos los ingredientes básicos de la EDC: el campo complejo del electrón, su complejo conjugado del campo del positrón, y el campo de conexión de gauge que nos da el electromagnetismo y los fotones. El paso final en el ensamblaje de la teoría es especificar cómo interaccionan entre sí los fotones y los electrones/positrones. De vuelta a la ecuación (5.6), la escribimos como simple producto de los tres campos (electrón, positrón y fotón). Pero en aquellos tiempos más sencillos no conocíamos la invariancia de gauge ni sabíamos que el campo del positrón es el complejo conjugado del campo del electrón. Básicamente íbamos por buen camino, pero ahora tenemos que ser un poco más cuidadosos.

Queremos que nuestra interacción sea invariante de gauge, teniendo en cuenta las propiedades de transformación del electrón (9.3) y del fotón (9.7). Podemos cancelar el $e^{i\theta}$ del campo del electrón ψ_e utilizándolo siempre en combinación multiplicativa con el campo del positrón $\psi_e{}^*$. Pero ese fotón parece que nos va a dar problemas. Una transformación de gauge añade $\partial_\mu\theta$ a A_μ, y no está claro qué puede cancelar eso.

Espere un momento, quizá esté claro. ¿Recuerda (9.6), la versión invariante de gauge de «mantener constante el campo del electrón»?

La expresión del lado izquierdo solamente recoge una fase global $e^{i\theta}$ cuando hacemos una transformación de gauge. Hay piezas adicionales molestas que se introducirían para cualquiera de los términos de forma individual, pero cuando se ponen juntas todas ellas se cancelan. Quizá podamos aprovecharnos de ello.

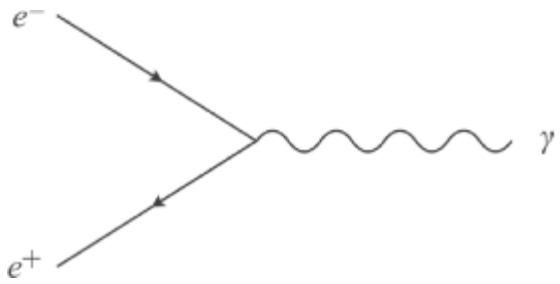
Para simplificar nuestra notación y mantener nuestra mente en la física esencial, vamos a ocultar algunos índices, escribiendo ∂ para ∂_μ y A para A_μ. Entonces el lado izquierdo de (9.6) es $\partial\psi_e - iA\psi_e$. No estamos realmente haciendo esto igual a 0: estamos simplemente aprovechando el hecho de que lo único que le ocurre bajo una transformación de gauge es que incorpora una fase global $e^{i\theta}$. Hay una forma obvia de deshacerse incluso de esa fase: multiplicar por el campo complejo conjugado (del positrón), que se transforma con la fase opuesta, ya que $\psi_e^* \to e^{-i\theta}\psi_e^*$. De este modo, podemos construir una combinación completamente invariante de gauge que sea elegible para aparecer en el lagrangiano:

$$\psi_e^*\left(\partial\psi_e - iA\psi_e\right) = \psi_e^*\,\partial\psi_e - i\psi_e^*\psi_e A\,. \qquad (9.16)$$

¡Intrigante! El segundo término se parece básicamente a la interacción que describimos en el capítulo 5: un producto de los campos del electrón, el positrón y el fotón, con una notación ligeramente diferente. Pero ¿qué es ese primer término, con solamente el positrón y una derivada del electrón? Fácil: es el término cinético. A diferencia de los campos escalares o del campo del fotón, el término cinético para los electrones (y otras partículas de espín ½) requiere una sola derivada en lugar de dos derivadas.

De este modo se revela otro secreto: la forma de la interacción entre fotones y electrones/positrones en EDC puede considerarse como el resultado de empezar con el término cinético, exigirle que sea invariante de gauge y darnos cuenta de que tenemos que compensarlo de un modo apropiado incluyendo un término con la conexión, también conocido como campo de gauge. El resultado difiere de (5.6)

en un factor numérico global, pero eso es solamente porque hemos sido bastante informales acerca de los factores numéricos globales. El segundo término en (9.16) es realmente el lagrangiano de interacción de la EDC, completamente determinado por los requisitos de invariancia de gauge. Esta bonita cualidad es válida para cualquier teoría de campo de gauge, no solamente para la EDC.



También podemos ver, entre bastidores, el teorema de Noether en acción. Este teorema nos dice que las simetrías continuas implican magnitudes conservadas. La simetría U(1) del electromagnetismo ciertamente cuenta, y la cantidad asociada es la carga eléctrica. Y ahora vemos cómo la invariancia de gauge lo hace posible. Para obtener un lagrangiano invariante de gauge, necesitamos incluir el campo del positrón, así como el campo del electrón en el término de interacción $\psi_e^* \psi_e A$. Lo que significa que el correspondiente vértice del diagrama de Feynman tendrá un electrón y un positrón entrantes, con un fotón saliente. Una carga –1 y una carga +1 entrantes y una carga 0 saliente, conservando la carga globalmente. Nunca podremos tener un vértice con dos electrones entrando y un solo fotón saliendo; la invariancia de gauge no lo permitirá.

---

1 Recuerde la regla de la derivada de un producto: $\partial_\mu (fg) = (\partial_\mu f)g + f(\partial_\mu g)$.

DIEZ

FASES

La electrodinámica cuántica tiene una historia fascinante llena de capítulos. Pero en cierto sentido su éxito no fue sorprendente. Sabemos que la mecánica cuántica funciona y sabemos que existe el electromagnetismo, así que tenían que encontrarse de alguna manera. Al principio, la renormalización supuso cierta confusión, pero afortunadamente todo salió bien.

El verdadero triunfo de la teoría cuántica de campos, que en aquel momento no parecía en absoluto inevitable, consistió en explicar las demás fuerzas de la naturaleza, en particular, las fuerzas nucleares fuerte y débil. La idea de que estas podrían ser algún tipo de generalizaciones de la EDC, esta vez basadas en simetrías de gauge no abelianas (no conmutativas), fue un atractivo que comenzó con Chen Ning Yang y Robert Mills en 1954, por lo que estos modelos se conocen como **teorías de Yang-Mills**. Pero hubo una serie de obstáculos inmediatos, y más de un físico se mostró escéptico de que el programa llegara a funcionar. Su éxito final se debió a una mejor comprensión de la riqueza de la teoría de campo de gauge no abeliana; en particular, del hecho de que las teorías pueden manifestarse en **fases** diferentes.

La palabra «fase» está tomada de la física macroscópica, donde se refiere a la forma en que un mismo tipo de materia puede aparecer en diferentes estados, con sus correspondientes propiedades físicas diferentes. El agua, por ejemplo, puede aparecer en forma de sólido (hielo), líquido o vapor gaseoso, en función de su temperatura y presión. Las distintas fases de un material tendrán características

diferentes, como la densidad o la velocidad del sonido en su interior, a pesar de estar hechas de la misma materia subyacente.

Las fases de las teorías de campo de gauge funcionan de forma análoga: un conjunto similar de ingredientes subyacentes puede mostrar distintos tipos de comportamiento observado. Teorías como el electromagnetismo y la relatividad general se encuentran en la **fase de Coulomb**, caracterizada por fuerzas de largo alcance que obedecen a una ley cuadrática inversa. En el lenguaje de las partículas diríamos que los bosones de gauge no tienen masa e interaccionan débilmente, lo que es cierto tanto para los fotones como para los gravitones. Una teoría podría estar en la **fase confinada** si los bosones de gauge siguen sin tener masa, pero interaccionan fuertemente y, por tanto, están capturados dentro de partículas compuestas. Es el caso de los gluones en las interacciones fuertes (CDC). Y por último tenemos la **fase de Higgs**, en la que la simetría de gauge subyacente se rompe espontáneamente. Entonces los bosones de gauge se vuelven masivos y la fuerza se vuelve de corto alcance, como en las interacciones débiles. Los pioneros de la teoría de campo de gauge no habían previsto esta increíble riqueza de estructuras; las ecuaciones, como siempre, son más inteligentes que nosotros.

CDC

En el capítulo anterior cantamos las alabanzas de la invariancia de gauge, una herramienta enormemente útil a la hora de entender la dinámica y las interacciones de los campos cuánticos. Una de las grandes implicaciones es que los bosones de gauge, como los fotones y los gravitones, no deben tener masa, razón por la cual se manifiestan como fuerzas de largo alcance.

Un momento. Ya hemos anticipado que las otras fuerzas de la naturaleza, las fuerzas nucleares fuerte y débil, también se describen mediante teorías de campo de gauge. Pero esas fuerzas definitivamente

no son de largo alcance. Un alcance típico asociado a la fuerza fuerte sería la longitud de onda Compton de un protón. Así que busque la masa del protón en electrón-voltios, conviértala en centímetros inversos, y tome el recíproco para obtener una distancia. Terminamos con un número del orden de 10^{-14} cm. Para la fuerza débil, haga lo mismo, pero con el bosón *W* en lugar del protón, para obtener 10^{-16} cm. En efecto, alcances muy cortos. ¿Qué está pasando?

Tanto en el caso de la fuerza fuerte como en el de la fuerza débil, existen razones sutiles y fascinantes por las que las fuerzas son de corto alcance, aunque se basen en simetrías de gauge. Y como al universo le gusta hacer que las cosas sean interesantes, las razones son completamente diferentes en los dos casos.

Pensemos primero en la fuerza nuclear fuerte: la cromodinámica cuántica, o CDC, responsable de la unión de quarks y gluones para formar partículas compuestas como los bariones y mesones. (La fuerza que mantiene unidos a los protones y neutrones en los núcleos puede considerarse un efecto secundario de lo que ocurre en el interior de los propios nucleones). Hay dos características que hacen que la CDC sea tan diferente en la práctica de la EDC: los gluones interaccionan directamente con otros gluones, y la renormalización hace que la interacción sea más fuerte, en lugar de más débil, a energías más bajas.

La CDC se basa en una simetría SU(3) a partir de campos de quarks en rotación en un espacio de color complejo tridimensional. Es una historia larga y algo tortuosa. La idea de generalizar de las simetrías de gauge abelianas a las no abelianas surgió de Yang y Mills en 1954, pero no sabían nada de los quarks. En su lugar, intentaban trabajar directamente con protones y neutrones, y ninguno de los intentos resultó realmente. No fue hasta 1964 cuando Murray Gell-Mann y George Zweig propusieron de forma independiente la idea de los quarks. Poco después, Otto Greenberg y, por separado, Moo-Young Han y Yoichiro Nambu propusieron las teorías de campo de gauge SU(3) con quarks como constituyentes fundamentales, pero los detalles de sus modelos eran muy diferentes de la actual imagen de la

CDC. Fueron Gell-Mann y Harald Fritzsch quienes introdujeron las cargas de color tal y como las entendemos ahora en 1971, y en publicaciones independientes en colaboración con William Bardeen y Heinrich Leutwyler, elaboraron la teoría tal y como la conocemos ahora (sin apreciar aún algunas de sus propiedades cruciales). Gell-Mann, que tenía ingenio para la terminología, bautizó la teoría como «cromodinámica cuántica» por analogía con la electrodinámica cuántica.

En muchos aspectos, la estructura básica de la CDC es muy parecida a la de la EDC. En ambos casos tenemos un campo de gauge de conexión, que da lugar a los fotones en la EDC y a los gluones en la CDC. En la EDC tenemos carga eléctrica, que es solamente un número, mientras que en la CDC tenemos cargas de color que existen en un espacio vectorial tridimensional rojo/verde/azul. En ambos casos, las partículas portadoras de fuerzas carecen de masa, en última instancia debido a las restricciones impuestas por la invariancia de gauge. La forma de los términos cinéticos y de las interacciones son en gran medida las mismas.

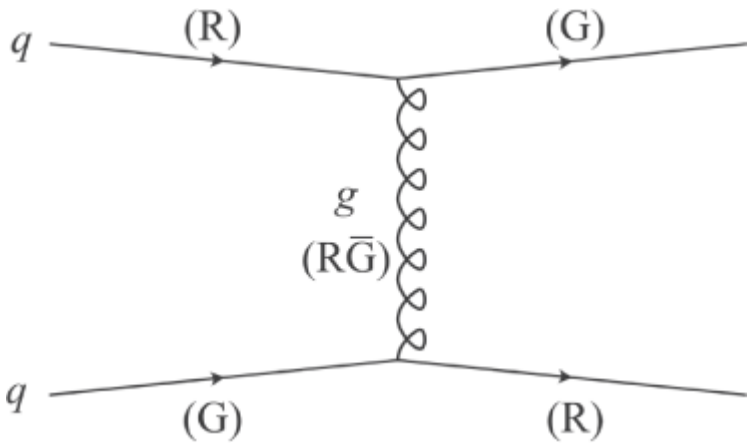
Hay una diferencia importante: SU(3) es un grupo no abeliano, mientras que U(1) es un grupo abeliano. Una transformación U(1) es simplemente una rotación en el plano complejo, o equivalentemente, una multiplicación por un factor de fase, $R = e^{i\theta}$. Así que no importa en qué orden las apliquemos: $R_1 R_2 = e^{i(\theta_1+\theta_2)} = R_2 R_1$. Pero una transformación SU(3) M es una matriz 3×3 y el orden en que apliquemos dos de ellas importa:

$$M_1 M_2 \neq M_2 M_1. \qquad (10.1)$$

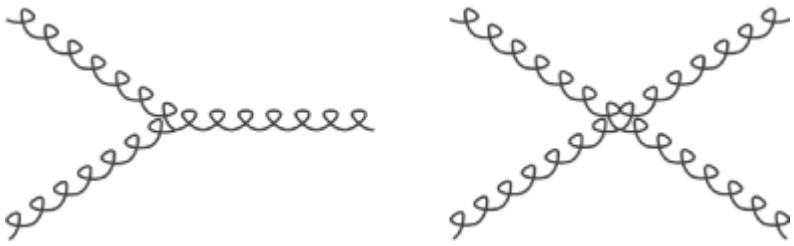
Parece un detalle matemático, pero tiene enormes consecuencias físicas. El hecho de que U(1) sea abeliano significa que las transformaciones sucesivas «pasan unas a través de otras» y, mediante un poco de matemáticas (que estamos ocultando), esto se traduce directamente en el hecho de que los fotones físicos pasan de largo sin

interaccionar. Las transformaciones SU(3) no pasan unas a través de otras, y los gluones, por su parte, tienden a chocar entre sí.

Más formalmente, a nivel de interacciones directas —vértices fundamentales del diagrama de Feynman que leeríamos a partir del lagrangiano básico— los fotones interaccionan solamente con partículas cargadas eléctricamente. Y el fotón en sí es neutro, por lo que los fotones no interaccionan directamente entre ellos mismos. (Hay interacciones efectivas de cuatro fotones una vez que imponemos un corte, pero en general son bastante débiles, y de todos modos sabemos que en realidad están inducidas por diagramas de bucles con partículas cargadas). En cambio, los propios gluones transportan una forma de carga de color. Esencialmente, cada gluon transporta tanto un color (rojo, verde o azul) como un *anticolor*. Así, un quark rojo y un quark verde pueden interaccionar intercambiando un gluon rojo/*antiverde*, cambiando así sus propios colores en el proceso.[1]



Este diagrama de intercambio de gluones no es tan diferente del caso EDC por sí mismo. Pero también hay vértices de tres y cuatro gluones como parte de los ingredientes fundamentales de los diagramas de Feynman de la CDC. La existencia de estos vértices es una consecuencia de la naturaleza no abeliana de la simetría de gauge SU(3).



CONFINAMIENTO

El otro ingrediente que desempeña un papel definitorio en las interacciones fuertes proviene de nuestra discusión del capítulo 6 sobre el grupo de renormalización y las constantes de acoplamiento variables. Allí explicamos que la constante de estructura fina efectiva de la EDC aumenta a medida que consideramos energías cada vez más altas. No fue hasta la década de 1970 cuando los físicos se dieron cuenta, para su sorpresa, de que en ciertas teorías de campo de gauge no abelianas el acoplamiento funciona de forma opuesta: parece mayor a energías bajas y menor a energías altas. El cálculo fue realizado por David Politzer, David Gross y Frank Wilczek en 1973, por el cual compartieron el Premio Nobel de Física de 2004. Esta propiedad se denomina **libertad asintótica**, ya que a medida que las energías de interacción aumentan asintóticamente, el acoplamiento tiende a 0: los quarks y los gluones actúan como partículas libres. Pero, por otro lado, a medida que las energías de interacción disminuyen —lo que, recordemos, corresponde a distancias mayores— el acoplamiento aumenta. En principio, podría hacerse infinitamente grande.

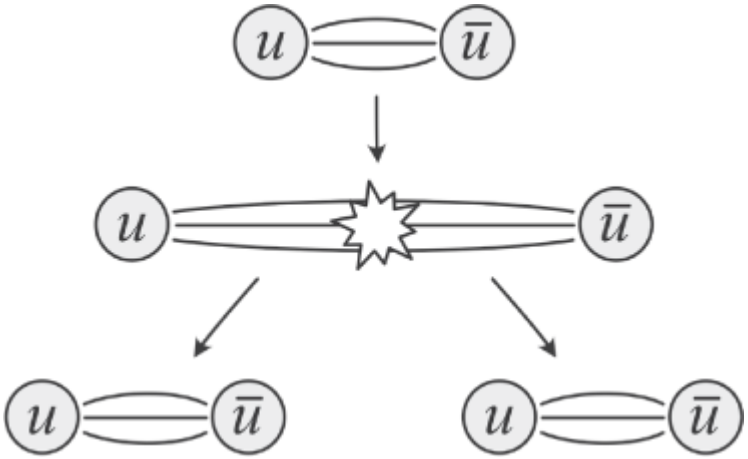
En la práctica, lo que esto significa es que los quarks y gluones de color en realidad nunca llegan a grandes distancias y bajas energías de interacción. Más bien tenemos el fenómeno del **confinamiento**: las partículas de color siempre están unidas a otras partículas de color en combinaciones sin color. El confinamiento explica cómo los gluones pueden carecer de masa y, sin embargo, la fuerza nuclear fuerte tiene un alcance tan corto: los gluones se mueven a la velocidad de la luz,

pero no viajan muy lejos. Siguen interaccionando con quarks y otros gluones: si amenazan con alejarse demasiado de sus compañeros, son irresistiblemente atraídos de vuelta.

Por supuesto, a estas alturas ya sabe que el párrafo anterior no es más que una historia evocadora. Lo que realmente ocurre es que hay campos cuánticos que interaccionan entre sí. En la CDC, los campos básicos con los que empezamos corresponden a quarks y gluones, pero esos no son los estados individuales similares a partículas con los que terminamos. Más bien, los campos de quarks y gluones se asientan en ciertas partículas colectivas que llamamos **hadrones**. Los quarks y los gluones dentro de un hadrón no son partículas literalmente puntuales que se mueven constantemente y chocan entre sí; por lo general, los campos son completamente estacionarios, o casi. Pero no hay nada malo en contar una historia evocadora siempre que comprendamos sus limitaciones.

Quizá piense: «Bueno, ¿y si intento separar los quarks entre sí? ¿Qué saldría mal?». Esto nos lleva a otra historia evocadora. Empieza por reconocer que el campo de gluones que interacciona fuertemente alrededor de un quark no se extiende uniformemente en todas direcciones como el campo eléctrico de una partícula cargada. Más bien, queda confinado en un **tubo de flujo** que se extiende a algún otro quark. Consideremos un quark *up* y un quark *antiup*, que pueden combinarse para formar (al menos durante un tiempo) una partícula llamada **pion neutro**. Si tomáramos unas pinzas realmente minúsculas y tratáramos de separar los dos quarks entre sí, estaríamos estirando el tubo de flujo. Eso requiere energía, ya que el tubo tiene una cantidad constante de energía por unidad de longitud. Al final, hemos introducido tanta energía en el sistema que resulta más favorable formar otro par quark/antiquark. Este par puede aparecer espontáneamente a lo largo del tubo de flujo, que entonces se divide en dos tubos de flujo. Es muy parecido a los extremos de una cuerda: si se corta una cuerda que inicialmente tiene dos extremos, no se obtienen

dos extremos separados, sino dos trozos de cuerda con dos extremos cada uno.



Antes de que conociéramos los gluones o los tubos de flujo que interaccionan con fuerza, se observó que ciertos hadrones seguían los patrones que cabría esperar de los modos de vibración de una cuerda. Esto llevó a la invención de algo llamado **teoría de cuerdas**. La idea nunca encajó realmente como teoría simplemente de interacciones fuertes, pero se convirtió en un marco para la gravedad cuántica y todas las demás interacciones juntas, y sigue siendo muy popular entre los físicos teóricos.

RUPTURA DE SIMETRÍA

Esto engloba prácticamente todo sobre las interacciones electromagnética y fuerte. Ambas fuerzas se basan en simetrías de gauge, y los bosones de gauge correspondientes carecen de masa. Algo parecido ocurre con la gravitación, descrita por la relatividad general. Los gravitones interaccionan entre sí, pero muy débilmente, por lo que no están confinados y la gravedad se encuentra en la fase de Coulomb.

Por último, pasamos a las interacciones débiles, en las que hay tres bosones de gauge: los cargados W^+ y W^- y el neutro Z_0. No carecen de

masa: los W tienen unos 80 GeV, y el Z unos 91 GeV. ¿Cómo pueden ser masivos los bosones portadores de fuerza si la invariancia de gauge nos prohíbe escribir algo como $m^2(A_\mu)^2$ en el lagrangiano? La respuesta es: porque la simetría subyacente se rompe de modo espontáneo.

Para simplificar las cosas, consideremos un modelo de juguete con un grupo de simetrías SO(2). Como grupo, SO(2) es lo mismo que U(1), pero vamos a pensar en él como rotaciones en un espacio vectorial real de dos dimensiones, en lugar de una sola dimensión compleja. Digamos que tenemos un campo escalar Φ que es un vector en este espacio interno bidimensional. Esto significa que tiene componentes Φ_1 y Φ_2, y todo el campo se puede escribir como

$$\Phi = \Phi_1 \vec{e}_1 + \Phi_2 \vec{e}_2 \,, \tag{10.2}$$

donde $\vec{e}_1$ y $\vec{e}_2$ son vectores de base. La simetría SO(2) rota el vector Φ dentro de este espacio.

Para hacer un lagrangiano, necesitamos construir cantidades que sean invariantes bajo la simetría. Hay un ejemplo obvio, el campo al cuadrado:

$$|\Phi|^2 = (\Phi_1)^2 + (\Phi_2)^2. \tag{10.3}$$

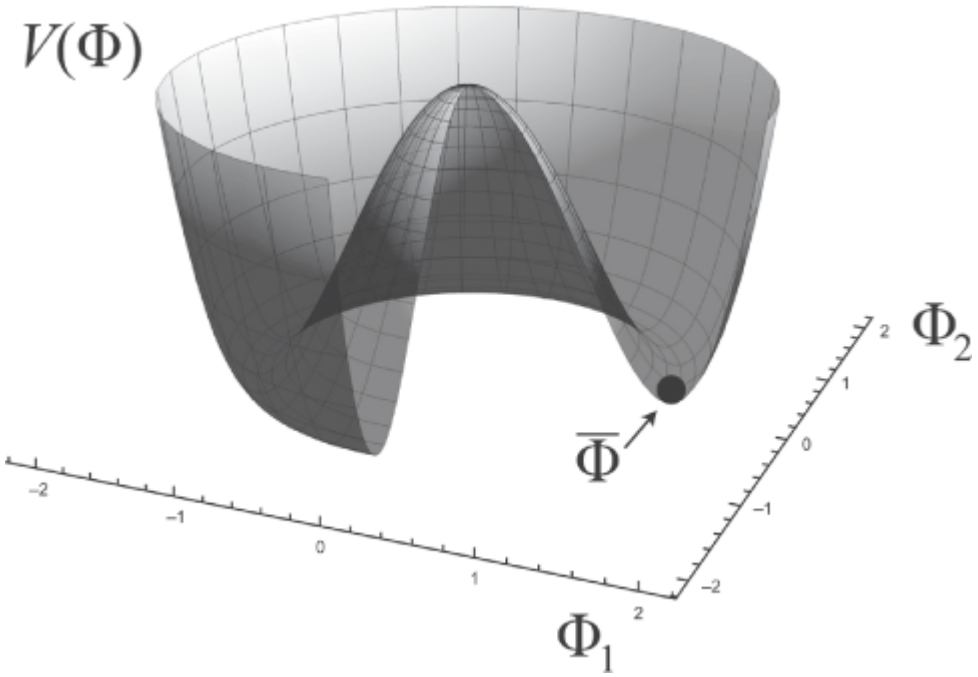
Esto no es más que la longitud al cuadrado del vector Φ, que no cambia cuando rotamos cualquier ángulo alrededor del origen. Así que es fácil hacer un término de masa invariante en SO(2), $m_\Phi^2|\Phi|^2$. Del mismo modo, podemos hacer un término de interacción simplemente elevando de nuevo al cuadrado para obtener $|\Phi|^4$. Puesto que $|\Phi|^2$ es invariante bajo la simetría, cualquier función de ella también lo será.

Consideremos el siguiente potencial para Φ:

$$V(\Phi) = -\mu^2 |\Phi|^2 + \lambda |\Phi|^4 \,. \tag{10.4}$$

Esto se conoce como «**potencial de sombrero mexicano**», por razones que son evidentes cuando lo representamos gráficamente, como en la página siguiente en una vista de corte. El parámetro μ tiene dimensiones de masa, y l es adimensional (como podemos deducir, porque un campo escalar como Φ tiene dimensiones de masa, y el potencial es parte de la densidad de Lagrange, que tiene dimensiones de $[E]^4$, que es lo mismo que $[M]^4$).

Lo interesante es que tenemos un signo menos delante del término $\mu^2 |\Phi|^2$. Este no era necesario, el signo podría ser más o menos, y solamente estamos explorando esta posibilidad. Como resultado, el mínimo del potencial no es el origen en $\Phi = 0$. Cerca del origen, el término $|\Phi|^2$ cobra importancia frente al término $|\Phi|^4$, y ese término tiene un coeficiente negativo. Así pues, el potencial disminuye a medida que nos alejamos del origen, para volver a aumentar una vez que el término $|\Phi|^4$ empieza a dominar.



Este comportamiento no tiene nada de extraño o erróneo. No existe ninguna regla que diga que el mínimo del potencial tenga que estar en el origen. Hay una regla que dice que tiene que *existir* un

mínimo en alguna parte. De lo contrario, no existiría un estado de vacío estable, pues el potencial campo decaería eternamente. Eso podría ser un universo alternativo interesante, pero no parece ser el mundo en el que vivimos.

Como solemos pensar que vivimos cerca del estado de vacío (energía mínima), el campo no va a estar centrado alrededor de 0; caerá hacia el pliegue del ala del sombrero. El potencial de sombrero mexicano es rotacionalmente simétrico respecto al origen, lo que refleja la simetría SO(2) subyacente. Así que no podemos decir de antemano dónde caerá el campo: todos los puntos del pliegue son equivalentes entre sí. Pero caerá en algún lugar. Denominamos $\overline{\Phi}$ al valor donde cae, llamado **valor de expectación de vacío** del campo.[2] («Valor de expectación» en lugar de simplemente «valor» porque estamos haciendo mecánica cuántica, después de todo. Si se midiera de manera repetida el valor del campo, se obtendrían respuestas ligeramente diferentes, pero su promedio sería $\overline{\Phi}$).

Toda la física subyacente de este campo escalar es completamente invariante con respecto a las transformaciones SO(2). Pero el lugar específico $\overline{\Phi}$ donde termina el campo *no* es invariante; una rotación lo desplaza a lo largo del pliegue del ala del sombrero. Al caer aleatoriamente hacia un punto concreto del pliegue, el valor del campo vulnera ahora la simetría subyacente. Por eso este fenómeno se llama **ruptura espontánea de la simetría**, porque la física subyacente es invariante, pero el valor de expectación específico de vacío no lo es.

Todo lo que hemos dicho hasta ahora sobre la ruptura espontánea de la simetría se aplicaría igualmente bien si la simetría fuera global (solamente se puede hacer una rotación uniforme en todo el espacio-tiempo) en lugar de gauge (se pueden especificar rotaciones diferentes en cada punto). La simetría global fue estudiada por primera vez en la década de 1960 por Yoichiro Nambu, Jeffrey Goldstone y otros. En ese caso ocurre algo interesante: debido a la simetría subyacente, siempre hay una dirección en el espacio del campo en la que podemos empezar en el valor de expectación de vacío y movernos de forma que

el potencial permanezca constante. (En nuestra imagen, es el pliegue del sombrero). Un potencial constante, incluso solo a lo largo de ciertas direcciones en el espacio del campo, corresponde a una partícula con masa cero, llamada bosón de Goldstone (o bosón de Nambu-Goldstone). Resulta que hay un teorema que se puede demostrar, y es que siempre que se rompe espontáneamente una simetría global exacta, solamente quedan una o más partículas escalares sin masa. Esto suena bien, pero también es problemático si se intenta describir el mundo real, ya que no se sabe que existan tales partículas.

Por fortuna, la ruptura espontánea de una simetría de gauge es un berenjenal completamente distinto. Consideremos ese caso.

EL MECANISMO DE HIGGS

En la década de 1960, los físicos aún no habían descifrado la libertad asintótica y el confinamiento, pero sabían perfectamente que las interacciones fuerte y nuclear débil eran de corto alcance. Buscaban una forma de conciliar la propuesta de Yang-Mills de una teoría de campo de gauge no abeliana con el problema de que los bosones de gauge no tuvieran masa. Parecía prometedor considerar la ruptura espontánea de simetría, que había demostrado desempeñar un papel crucial en la comprensión de la superconductividad. Pero existía el problema de los bosones de Goldstone sin masa, que nadie había observado nunca.

El hecho de que la ruptura espontánea de la simetría funcione de forma muy distinta para las simetrías de gauge y las simetrías globales fue señalado por primera vez por el físico de la materia condensada Philip Anderson y aplicado en el contexto de la física de partículas por tres grupos en 1964: Robert Brout y François Englert; Peter Higgs; y Gerald Guralnik, Carl Hagen y Tom Kibble. La idea básica es que dos características problemáticas —los bosones de gauge sin masa y los

bosones de Goldstone sin masa— se anulan mutuamente. Cuando se rompe de modo espontáneo una simetría de gauge, los bosones de gauge asociados se «comen» a los bosones de Goldstone. Los bosones de Goldstone desaparecen como partículas independientes y los bosones de gauge se vuelven masivos en lugar de no tener masa. Es conveniente denominar lo que ocurre simplemente como el mecanismo de Higgs, pero todos estos colegas merecen algo de reconocimiento. Englert y Higgs compartieron el Premio Nobel en 2013; Brout ya había fallecido, y el Nobel no se concede a título póstumo. (Se tarda casi medio siglo entre proponer una idea y ganar el Nobel por ella, si se hace la cuenta).

Ahora es el momento de sacar provecho del duro trabajo que hicimos para entender cómo hacer lagrangianos invariantes de gauge y utilizarlo para ver explícitamente cómo se vuelven masivos los bosones de gauge. El truco está en el lagrangiano cinético para el campo escalar, que va a obtener un valor de expectación de vacío en un potencial de sombrero mexicano. Pensemos en nuestro ejemplo de SO(2) con el escalar Φ, y ahora estamos considerando explícitamente una simetría de gauge, por lo que también tenemos un campo de conexión A_μ.

En una teoría ordinaria de campo escalar, el término cinético sería algo así como $(\partial_\mu \Phi)^2$, pero este por sí mismo no es invariante de gauge. En lugar de esto, volvamos a (9.6), donde hicimos una combinación invariante de gauge de la derivada del campo del electrón y la conexión multiplicada por ese campo. Ahora estamos pensando en nuestro escalar Φ en lugar del electrón, y estamos pensando en SO(2) y no en U(1), por lo que no hay números imaginarios, pero las ideas básicas son las mismas. Un posible término cinético invariante de gauge es[3]

$$|\partial_\mu \Phi - A_\mu \Phi|^2 = (\partial_\mu \Phi)^2 + (A_\mu)^2 |\Phi|^2. \qquad (10.5)$$

Parece que, en el lado derecho, la primera pieza es un término cinético ordinario y la segunda es una interacción que implica dos bosones de gauge y dos partículas Φ. Y de hecho sería así, si no fuera por la ruptura espontánea de la simetría.

Cuando hablamos de «partículas» en la teoría cuántica de campos, estamos pensando en cuantos que surgen como excitaciones en el campo, pequeñas perturbaciones por encima del estado de vacío. Pero el estado de vacío aquí no es $\Phi = 0$, es $\Phi = \overline{\Phi}$. Así que para revelar qué tipo de partículas predice esta teoría, es útil cambiar las variables:

$$\Phi(x, t) = \overline{\Phi} + \phi(x, t). \tag{10.6}$$

Aquí $\overline{\Phi}$ es el valor de expectación de vacío, que no es más que una constante, y $\phi(x, t)$ es un campo que nos dice lo lejos que estamos de $\overline{\Phi}$. No tenemos que hacer esta sustitución, es solamente por comodidad, ya que ahora el valor de vacío de nuestro campo es $\phi = 0$. Es el campo desplazado ϕ, no el campo original Φ, cuyas excitaciones se interpretan naturalmente como partículas.

Veamos qué ocurre cuando introducimos el cambio de variables (10.6) en la segunda parte de «interacción» del término cinético (10.4):

$$\left(A_\mu\right)^2 |\Phi|^2 = \overline{\Phi}^2 \left(A_\mu\right)^2 + 2\overline{\Phi}\phi\left(A_\mu\right)^2 + \phi^2\left(A_\mu\right)^2. \tag{10.7}$$

El último término aquí es solamente la interacción que hemos discutido anteriormente, entre dos escalares y dos bosones de gauge. Para interpretar el penúltimo término, recuerde que $\overline{\Phi}$ es solamente un número constante, no un campo dinámico. Así que este término es otra interacción, entre un único f y dos bosones de gauge, con $2\overline{\Phi}$ actuando como constante de acoplamiento.

Sin embargo, el primer término del lado derecho amenaza con implicaciones. $\overline{\Phi}$ es solamente una constante, lo que significa que el

único contenido de campo es $(A_\mu)^2$. ¡Así que, en realidad, se trata de un término de masa para el bosón de gauge! Se suponía que la invariancia de gauge evitaría que eso sucediera. ¿Qué está pasando?

En un sentido legítimo, lo que sucede son las manipulaciones matemáticas por las que acabamos de pasar. En una teoría de campo de gauge, no podemos simplemente introducir un término de masa para el campo de gauge en nuestro lagrangiano sin vulnerar la invariancia de gauge. Pero la ruptura espontánea de la simetría provoca la existencia de un término de masa efectiva, donde el papel de la masa lo desempeña el valor de expectación de vacío del campo que rompe la simetría. (Si hubiéramos sido más cuidadosos, también habría habido constantes de acoplamiento que relacionan $\overline{\Phi}$ con la masa de A_μ, pero aquí vamos a conceptos generales más que a fórmulas precisas). Aunque no podamos verlo explícitamente a partir de lo que hemos hecho, el bosón de Goldstone sin masa también ha desaparecido, al haber sido devorado por el bosón de gauge.

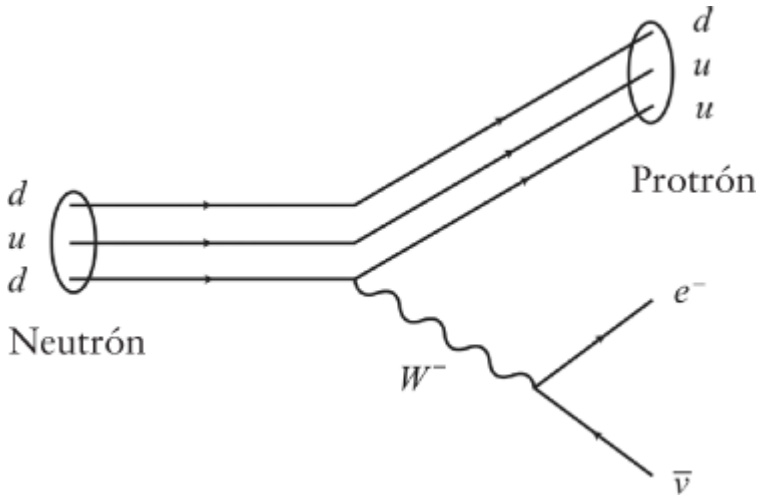
Pero seguramente se sentiría decepcionado si no hubiera una historia evocadora que contar. He aquí una forma de verlo: un campo de gauge sin masa, no asintóticamente libre, como el que tenemos en el electromagnetismo o la gravedad, da lugar a una ley de fuerza inversa cuadrática de largo alcance, básicamente porque las líneas de fuerza viajan eternamente y se diluyen con la distancia, como se explica en *Espacio, tiempo y movimiento*. Ahora tenemos un nuevo campo Φ que impregna el espacio, con un valor de expectación distinto de cero en todas partes. Y se trata de un campo cargado, en lo que respecta a nuestros bosones de gauge; se transforma bajo la transformación de gauge. Así que las líneas de fuerza no viajan libremente hasta el infinito. Viajan a través del campo escalar ambiente y son absorbidas gradualmente por él, como los rayos de luz que se desvanecen rápidamente en una atmósfera cargada de niebla. Por eso este campo de gauge es de corto alcance y no de largo alcance. Y eso, a su vez, corresponde a bosones de gauge masivos, en lugar de no masivos.

TEORÍA ELECTRODÉBIL

Los pioneros del mecanismo de Higgs se centraron en la fuerza nuclear fuerte, pero la idea acabó encontrando su lugar en nuestra comprensión moderna de la fuerza nuclear débil. El camino hasta llegar ahí fue, como es característico en la historia de la teoría cuántica de campos, tortuoso e impredecible.

La primera teoría bastante exitosa de las interacciones débiles fue la teoría de la desintegración beta de Enrico Fermi, de 1933, que describía la desintegración de un neutrón en un protón, un electrón y un antineutrino. Se trata de uno de los primeros ejemplos influyentes de teoría cuántica de campos y en ella tiene cabida el neutrino, que en aquella época aún era puramente hipotético. Además, no era renormalizable, lo que preocupaba a los físicos de la época. (Fermi postulaba una interacción directa entre cuatro fermiones, cuyos campos tienen dimensión $[E]^{3/2}$, por lo que su término de interacción tiene dimensión $[E]^6$, y es por tanto un operador irrelevante). Hoy en día diríamos que la teoría de Fermi es una teoría efectiva de baja energía perfectamente respetable para las interacciones débiles.

Una forma de construir una teoría renormalizable que se acerque a la de Fermi a bajas energías podría ser introducir un bosón cargado y masivo que pudiera mediar en la interacción, sustituyendo un vértice directo con cuatro fermiones por dos vértices, cada uno con dos fermiones, y el nuevo bosón. En el lenguaje moderno de los quarks y los diagramas de Feynman, un neutrón podría desintegrarse en protón cuando uno de sus quarks *down* (carga –⅓) pudiera convertirse en un quark *up* (carga +⅔) emitiendo un bosón W con carga –1, que posteriormente se convertiría en un electrón y un antineutrino.



Si se ha estado preguntando cómo distinguimos entre el neutrino electrónico, el neutrino muónico y el neutrino tau, así es: cuando el *W* se convierte en un leptón cargado y un antineutrino, siempre se trata del tipo de antineutrino el que está asociado al leptón que se produjo.

Pero como aún no se conocía el mecanismo de Higgs, era difícil construir una teoría con un bosón gauge masivo. En la década de 1950 Julian Schwinger hizo un intento, utilizando un modelo con simetría SU(2), pero su teoría tuvo la consecuencia no deseada de predecir también un bosón de gauge neutro sin masa. Con el tiempo se le ocurrió la idea de convertir ese error en una característica, interpretando el bosón neutro sin masa como el conocido fotón. De ahí nació la ambición de **unificar** las interacciones débil y electromagnética. Schwinger pasó el problema a su talentoso alumno Sheldon Glashow, quien sugirió un modelo con simetría tanto SU(2) como U(1). En el Reino Unido, Abdus Salam y John Ward exploraron modelos similares. Pero el mecanismo por el que se suponía que cualquiera de los bosones de gauge se volvía masivo seguía siendo un misterio a principios de la década de 1960. Fue Steven Weinberg, y también Salam en un trabajo no publicado, quienes introdujeron el mecanismo de Higgs de ruptura espontánea de la simetría de gauge en los intentos de unificar el electromagnetismo y la fuerza débil, dando lugar a lo que ahora llamamos la **teoría electrodébil**.

La teoría resultante es una construcción compleja. Comenzamos con una simetría de gauge SU(2) (un grupo tridimensional, por tanto, tres bosones de gauge) y una simetría de gauge U(1) separada (un bosón de gauge), así como un campo escalar complejo que se transforma como un vector bidimensional bajo SU(2), y que también está cargado bajo U(1). Cuando este campo —ahora conocido simplemente como campo de Higgs— adquiere un valor de expectación de vacío, tanto la simetría SU(2) como la U(1) se rompen de manera espontánea, pero una cierta combinación de ellas permanece, sin embargo, intacta y persiste como la U(1) del electromagnetismo. Existen tres bosones de gauge masivos —los bosones con carga W^+ y W^- y el bosón neutro Z^0— y el fotón, sin masa. Por tanto, aunque la teoría comience como SU(2) × U(1) y se descomponga en solamente U(1), la simetría de gauge final no es precisamente la parte U(1) original, sino una combinación de esta y parte de la SU(2).

Ya le dije que era complicado. Pero hay otra vuelta de tuerca que resulta ser tremendamente importante: en la década de 1950, siguiendo una sugerencia de los teóricos Tsung-Dao Lee y Chen Ning Yang, el físico experimental Chien-Shiung Wu demostró que las interacciones débiles vulneran la **paridad**, es decir, la transformación que intercambia una configuración física por su imagen especular. En la teoría electrodébil de Weinberg-Salam, los fermiones como electrones, neutrinos y quarks no tienen masa y se mueven, en consecuencia, a la velocidad de la luz. También son partículas de espín ½ y, por tanto, se caracterizan por una cierta **helicidad**, que describe si tienen espín hacia arriba o espín hacia abajo cuando se mide a lo largo del eje definido por su movimiento. (Si las partículas fueran masivas, siempre podríamos trabajar en su marco de reposo, donde no hay movimiento, pero si no tienen masa siempre se mueven a la velocidad de la luz hagamos lo que hagamos). La helicidad se clasifica en «a izquierdas» o levógira, y «a derechas» o dextrógira, lo que corresponde a que apuntemos con la mano izquierda o con la derecha

a lo largo de la dirección del movimiento mientras los dedos curvados indican el giro. Weinberg y Salam postularon que la simetría de gauge original SU(2) se aplicaba solo a los fermiones levógiros, dejando inalterados los dextrógiros. Esto es válido tanto para los quarks como para los leptones.

Lo cual es estupendo, salvo que, como ya hemos señalado, en el mundo real los electrones tienen masa. Afortunadamente, la ruptura espontánea de la simetría viene de nuevo al rescate. Para cualquier fermión ψ, podemos escribir una interacción invariante de gauge de la forma $y\Phi\psi^*\psi$. Hay tres campos implicados: ψ es un fermión levógiro, ψ es un antifermión dextrógiro y Φ es el campo escalar de Higgs. El factor numérico y es una constante llamada de **acoplamiento de Yukawa**, en honor al físico japonés Hideki Yukawa. El acoplamiento de Yukawa es un número adimensional, porque los fermiones tienen dimensión $[M]^{3/2}$, el Higgs es un campo escalar con dimensión [M], y cada término en la densidad de Lagrange tiene dimensión $[M]^4$. Habrá un término separado de esta clase para cada tipo de quark y leptón cargado.

Cuando el Higgs obtiene su valor de expectación de vacío, podemos escribir $\Phi \rightarrow \overline{\Phi} + H(x,t)$, donde $\overline{\Phi}$ es el valor de expectación fijo y H es el campo dinámico del bosón de Higgs. Veamos entonces qué ocurre con la interacción Higgs/fermión:

$$y\Phi\psi^*\psi \rightarrow y\overline{\Phi}\psi^*\psi + yH\psi^*\psi\,. \qquad (10.8)$$

Lo que era un único término se ha convertido ahora en dos. El segundo, $yH\ \psi^*\psi$, es solamente un término de interacción entre el Higgs y un par fermión/antifermión, y su intensidad es proporcional a y. En el primero, sin embargo, Φno es un campo dinámico, es una constante fija. Así que hay solamente dos campos implicados en él, el fermión y su antifermión. ¡Lo que significa que se trata simplemente de un término de masa para el fermión![4] El valor de expectación $\overline{\Phi}$ tiene dimensiones de masa, y el acoplamiento de Yukawa es

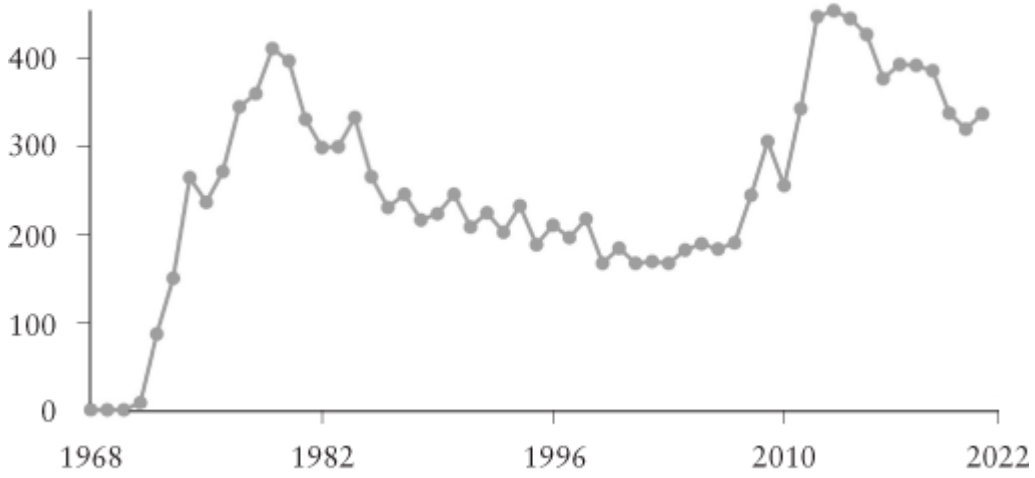
adimensional, por lo que las unidades funcionan correctamente para que la combinación $y\overline{\Phi}$ sea una masa. Cada fermión puede tener una masa diferente, ya que y será diferente para cada sabor de partícula, pero todas son proporcionales al valor de expectación del bosón de Higgs. Y como extra, esto arroja una predicción muy fuerte: que el acoplamiento entre el bosón de Higgs H y cada fermión es proporcional a la masa del fermión, ya que ambos son proporcionales a y. Esta característica fue útil a la hora de buscar el Higgs experimentalmente.

El hecho de que el Modelo Estándar trate de forma diferente a los fermiones levógiros y dextrógiros es la razón por la que, aunque Dirac no tuvo que pensar mucho para escribir un término de masa para el electrón en su ecuación original, hoy en día tenemos que «explicar» el origen de la masa. Dirac, en la década de 1930, no conocía la vulneración de la paridad, así que se limitó a escribir la masa. Hoy en día estamos lo suficientemente limitados como para tener que trabajar un poco más para explicar de dónde viene la masa del electrón, y en última instancia la atribuimos al campo de Higgs que impregna todo el espacio. No es que la noción misma de «masa» requiera un campo de Higgs, sino que la estructura restrictiva del Modelo Estándar prohibiría las masas de los fermiones si no fuera porque el bosón de Higgs obtiene un valor de expectación de vacío y rompe la simetría $SU(2)\times U(1)$. Esta hipótesis, concebida en la década de 1960, se ha confirmado espectacularmente en una larga serie de resultados experimentales de la década de 1970 hasta llegar al descubrimiento del bosón de Higgs en 2012.

Las ideas realmente buenas son escasas, y a veces es difícil incluso reconocer una cuando la ves. Cuando Weinberg publicó su artículo en 1967 con el modesto título «Un modelo de leptones», nadie le prestó mucha atención, ni siquiera el propio Weinberg. En la figura vemos el número de citas al artículo de Weinberg por año. No hubo ningún interés en los primeros años, luego aumenta a partir de 1971, se estabiliza un poco y vuelve a aumentar en 2010. Durante mucho

tiempo fue el artículo más citado de la física teórica. Sabemos lo que ocurrió alrededor de 2010: el Gran Colisionador de Hadrones inició seriamente la búsqueda del bosón de Higgs y acabó descubriéndolo en 2012. ¿Qué ocurrió en 1971?

S. Weinberg, «A Model of Leptons»,
Physical Review Letters (1967).
Citaciones por año, 1968-2022



https://inspirehep.net/literature/51188

Recordemos que en aquellos días la filosofía de las teorías de campo efectivo no había cuajado realmente, y se ponía mucho énfasis en si una teoría era renormalizable. Y la cosa era que, en el caso de la teoría de Weinberg, eso estaba lejos de ser obvio. El propio Weinberg intentó verificar si era renormalizable y se quedó atascado; otros ni siquiera lo intentaron, ya que se pensaba que los bosones de gauge masivos eran un signo de no renormalizabilidad.

Lo que ocurrió en 1971 fue que Gerard 't Hooft y su asesor de doctorado Martinus Veltman demostraron que la teoría es, de hecho, renormalizable. De repente, todos los físicos de partículas del mundo se interesaron por ella. Se trata de un caso interesante en el que los científicos vieron cómo aumentaba de manera drástica su confianza en

que una teoría determinada iba por buen camino, no por ningún resultado experimental impactante, sino simplemente por una mejor comprensión teórica.

La teoría electrodébil más el modelo CDC de interacciones fuertes tiene una simetría de gauge global de SU(3)×SU(2)×U(1). Junto con el conjunto específico de partículas fermiónicas (seis quarks, seis leptones) y, por supuesto, el campo de Higgs, la teoría general se conoce con el anodino nombre de **Modelo Estándar de la física de partículas**. Los últimos retoques teóricos del Modelo Estándar se añadieron en la década de 1970, y los físicos experimentales han seguido verificando la existencia de todas sus partículas. La versión original del Modelo Estándar fijaba todas las masas de los neutrinos en 0, pero no es difícil ampliar esto para permitir neutrinos masivos; el Premio Nobel de Física de 2015 fue concedido a Takaaki Kajita y Arthur McDonald por demostrar que los diferentes sabores de neutrinos pueden sufrir oscilaciones cuánticas de uno a otro, lo cual solo es posible si tienen masa. (Los valores exactos de las masas de los neutrinos todavía están pendientes de ser definidos). Aparte de esta pequeña modificación, el Modelo Estándar ha superado triunfalmente todas las pruebas experimentales que le hemos lanzado hasta la fecha.

---

1 Esta rápida historia podría llevarle a pensar que debería haber nueve tipos diferentes de gluones (tres colores por tres *anticolores*), pero una combinación global no contribuye, así que en realidad solamente hay ocho tipos de gluones. Eso es coherente con que SU(3) sea un grupo de ocho dimensiones ($3^2-1 = 8$). Siempre hay un tipo de bosón de gauge por dimensión del grupo de gauge.

2 Podría preguntarse si el campo podría caer a diferentes valores en diferentes puntos del espacio. Esto es complicado debido a la simetría de gauge, pero esencialmente sí, puede ocurrir. En general, ese es un estado de mayor energía que un campo constante, así que lo que ocurre es que el campo se endereza a sí mismo a medida que esa energía se convierte en otras partículas.

3 Estamos ocultando algún truco matemático que explica por qué no tenemos un término cruzado $-2A\ \Phi(\partial_\mu\Phi)$, pero créame que no pasa nada. Para profundizar más, tendríamos que

prestar más atención al hecho de que Φ es un vector de dos componentes y cada componente de A_μ es una matriz 2×2.

4 Me han dicho que es una buena práctica no poner demasiados signos de exclamación al escribir libros. Pero este es bien merecido.

ONCE

MATERIA

Los matices de la física de partículas pueden parecer algo distantes de las preocupaciones de nuestra vida cotidiana. Es fácil olvidar que todos llevamos dentro quarks, leptones y bosones de gauge. Estamos literalmente hechos de ellos. En estos dos últimos capítulos, iniciamos el viaje que conecta la física que llevamos dentro con el mundo que vemos habitualmente.

Hay algunas preguntas obvias: ¿por qué los átomos no son blandos? ¿Cómo es posible juntar una colección de átomos para formar un objeto sólido?

En el caso de un átomo único, no es demasiado difícil entender por qué mantiene su forma, pero tenemos que liberarnos de una idea errónea muy común. A veces oirá a la gente decir que «los átomos son en su mayoría espacio vacío». La gente no ha sido sincera con usted. Eso podría haber sido cierto si aún creyera en los modelos del átomo de Rutherford o de Bohr, con electrones puntiformes orbitando alrededor del núcleo como un pequeño sistema solar. En ese caso, la solidez de los átomos sería difícil de entender. Pero los átomos no son espacio vacío, al menos no en nuestra forma de pensar realista de la función de onda. La función de onda del electrón se extiende en una forma definida dentro de cada átomo. Se puede intentar deformar esa forma, pero generalmente costará mucha energía.

Eso ayuda a explicar la forma de un solo átomo, pero ¿cómo debemos entender el hecho de que varios átomos puedan *combinarse* para formar objetos sólidos? ¿Por qué no pueden simplemente apilarse los electrones de dos átomos?

No se debe, como a veces se cree, a la repulsión electromagnética de los electrones. En 1842, Samuel Earnshaw demostró un teorema según el cual no es posible atrapar partículas cargadas, como los electrones, en una configuración estable utilizando únicamente campos eléctricos estáticos, como los que existen en una colección de átomos. Como ocurre con la mayoría de los buenos teoremas, se puede intentar eludir las conclusiones negando algunos de los supuestos, pero la intuición va por buen camino. La solidez de la materia es algo más sutil.

La verdadera razón por la que la materia es sólida se reduce al hecho de que los electrones son fermiones, y los fermiones tienen una propiedad especial: no puede haber dos de ellos en el mismo estado cuántico. Por eso no se pueden apilar unos sobre otros. En este capítulo exploraremos esta característica y su conexión con el espín de las partículas implicadas.

BOSONES

¿Por qué todos los electrones tienen la misma masa y la misma carga? No es, como sugirió una vez John Wheeler a Richard Feynman, porque «todos son el mismo electrón». La idea de Wheeler surgió de la noción de que los positrones podrían considerarse como electrones que se mueven hacia atrás en el tiempo. Así, una interacción en la que un electrón y un positrón se aniquilan produciendo un fotón podría considerarse como un único electrón que entra y luego «rebota hacia atrás en el tiempo» para formar el positrón. En realidad, esta no es la forma correcta de pensar en los electrones, pero ayudó a Richard Feynman a inventar los diagramas de Feynman.

La verdadera razón por la que todos los electrones tienen la misma masa y carga es que todos son excitaciones de un único campo de electrones subyacente. Como resultado, dos electrones cualesquiera son **partículas idénticas**: son indistinguibles entre sí, incluso en

principio. Esto, sumado a la magia del entrelazamiento cuántico, tiene consecuencias cruciales para su comportamiento en la naturaleza.

Pensemos en partículas idénticas a nivel de funciones de onda. Supongamos que tenemos dos partículas idénticas, y utilizamos x_1 para denotar la variable de posición de la primera, y x_2 para la segunda. Entonces tenemos una función de onda global $\Psi(x_1, x_2)$. Ninguna de las partículas tiene «una posición»; las variables x_1 y x_2 son dos etiquetas diferentes en el mismo espacio tridimensional donde podrían observarse. Pero si observamos una de ellas, el hecho de que sean idénticas significa que no existe nada parecido a «cuál de ellas» observamos. Ninguna cantidad físicamente observable debería cambiar si intercambiáramos x_1 por x_2.

Hay una forma obvia de hacer que esto sea realidad: considerar funciones de onda que permanezcan literalmente inalteradas cuando cambiamos las dos variables de posición. Tales partículas existen: son los **bosones** de los que hemos estado hablando durante un tiempo, como los bosones de gauge y el bosón de Higgs. Llevan el nombre de Satyendra Nath Bose, que estudió por primera vez su comportamiento en la década de 1920. La función de onda de dos bosones es invariante al intercambiar las partículas entre sí:

$$\text{Bosones: } \Psi_B(x_1, x_2) = \Psi_B(x_2, x_1). \qquad (11.1)$$

En este caso, podemos pensar en dos funciones de onda independientes de una partícula, y_1 y y_2, y entrelazarlas de forma simétrica:

$$\Psi_B(x_1, x_2) = \psi_1(x_1)\,\psi_2(x_2) + \psi_1(x_2)\,\psi_2(x_1). \quad (11.2)$$

El mismo patrón se extiende a más de dos partículas de forma obvia.

No hay problema en tener dos bosones exactamente en el mismo estado cuántico. En (11.2) simplemente establecemos $\psi_1 = \psi_2$. De hecho, resulta que a los bosones les gusta estar en el mismo estado cuántico; si ya hay bosones en un estado determinado, aumenta la probabilidad de que otros bosones pasen a ese estado. Esto da lugar a importantes fenómenos físicos, como los láseres y los condensados, en los que un gran número de partículas comparten el mismo estado cuántico de baja energía. También permite la formación de campos de fuerza macroscópicos, como los de la gravedad y el electromagnetismo.

Bose, que trabajaba en la Universidad de Dhaka, en una parte de la India Británica que hoy es Bangladesh, escribió un artículo sobre las propiedades estadísticas de partículas idénticas y se lo envió a Albert Einstein, con la esperanza de que lo tradujera del inglés al alemán para su posible publicación en *Zeitschrift für Physik*, una de las revistas más importantes de la época. Era una petición un tanto descarada, pero no totalmente injustificada: Bose ya había traducido al inglés algunos de los artículos de Einstein para difundirlos en la India. Einstein estaba muy impresionado con el artículo de Bose; él mismo hizo la traducción, la envió a la revista con el nombre de Bose y escribió un artículo complementario que presentó a la misma revista. El comportamiento estadístico de los bosones, incluida su predilección por apilarse en el mismo estado cuántico, se conoce hoy como **estadística de Bose-Einstein**.

FERMIONES

Pero los electrones, con los que empezamos esta historia, no son bosones en absoluto. Esto es posible porque, en mecánica cuántica, el requisito de que el intercambio de partículas idénticas dé lugar a las mismas propiedades físicas no significa necesariamente que las funciones de onda deban permanecer simplemente inalteradas bajo un

intercambio de partículas, como en (11.1). También es posible que la función de onda adquiera un signo menos. Esto dejaría inalterada la probabilidad de observar la posición de las partículas, dada por la regla de Born como $P(x_1, x_2) = |\Psi(x_1, x_2)|^2$. Efectivamente, existen partículas de este tipo también; son los **fermiones**, llamados así en honor a Enrico Fermi.

$$\text{Fermiones: } \Psi_F(x_1, x_2) = -\Psi_F(x_2, x_1). \qquad (11.3)$$

Podemos hacer con los fermiones el mismo truco que hicimos con los bosones, construir la función de onda de dos partículas entrelazando dos funciones de onda de una partícula, pero esta vez en una combinación antisimétrica en lugar de simétrica:

$$\Psi_F(x_1, x_2) = \psi_1(x_1)\,\psi_2(x_2) - \psi_1(x_2)\,\psi_2(x_1). \qquad (11.4)$$

Las partículas de «materia» del Modelo Estándar son fermiones. Incluyen los seis sabores de quarks (*up, down, charm, strange, top, bottom*) y los seis sabores de leptones (electrón, muon, tau, neutrino electrónico, neutrino muónico y neutrino tau).

Pero ahora ocurre algo notable. Hemos mencionado que era fácil poner dos bosones en la misma función de onda y, de hecho, que preferían estar cerca de otros bosones de esa manera. Los fermiones son lo contrario. Si intentamos establecer $\psi_1 = \psi_2$, vemos inmediatamente que (11.4) desaparece: no es una función de onda en absoluto. Este es el famoso **principio de exclusión de Pauli**, postulado por primera vez por Wolfgang Pauli: dos fermiones no pueden ocupar exactamente el mismo estado cuántico.

En otras palabras, los fermiones ocupan espacio. Por eso se asocian generalmente con la «materia», a diferencia de las «fuerzas» formadas por bosones. Ahora sabemos que todos ellos son simplemente campos cuánticos que obedecen a estadísticas diferentes. Las propiedades de los fermiones se resumen en las denominadas **estadísticas de Fermi-**

Dirac, ya que Fermi y Dirac las estudiaron de forma independiente, pero Fermi llegó a sus conclusiones un poco antes.

Dado que el signo menos en (11.3) está permitido porque lo que realmente interesa es la probabilidad $|\Psi(x_1, x_2)|^2$, podría preguntarse si existen otros tipos de partículas cuyas funciones de onda se transformen como $\Psi(x_2, x_1) = e^{i\theta}\Psi(x_1, x_2)$ para algún parámetro fijo q, ya que esto también dejaría inalterada la probabilidad. El problema ahí es que la noción de «intercambiar dos partículas» tiene la propiedad de que si intercambiamos una vez y luego volvemos a intercambiar, queremos volver a la configuración original. Eso solamente ocurrirá si $e^{i\theta}$ es +1 o –1. En el caso especial en el que nos limitemos a dos dimensiones espaciales en lugar de las tres habituales, las reglas son un poco más laxas y se permite un nuevo tipo de partículas llamadas **anyones**. Jon Leinaas, junto a Jan Myrheim y sus colegas, formularon la hipótesis de su existencia en 1977, y Frank Wilczek dilucidó sus propiedades en 1982. Las partículas fundamentales del Modelo Estándar existen en tres dimensiones y son estrictamente bosones o fermiones, pero hay materiales que soportan excitaciones colectivas que existen efectivamente en solo dos dimensiones. La existencia de anyones en tales sistemas no se verificó experimentalmente hasta 2020.

ESPÍN Y ROTACIÓN

Para hacer hincapié, las propiedades de las funciones de onda multipartícula —o bien permanecen iguales bajo intercambio, o bien adquieren un signo menos— son las que definen a las partículas como bosones o fermiones. Pero a menudo se oye decir que los bosones son partículas cuyo espín intrínseco es un número entero —0, 1, 2, etcétera —, mientras que los fermiones son partículas con espín semientero —1/2, 3/2, 5/2, etcétera—. ¿Cuál es la relación entre estas dos ideas?

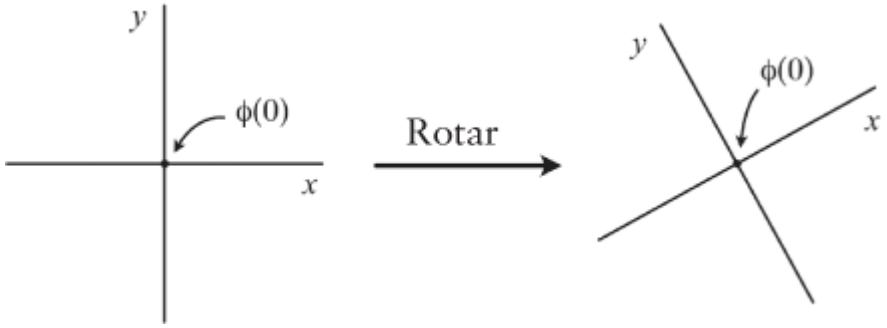
La relación es que podemos demostrar, en el marco de la teoría cuántica relativista de campos, que las partículas de espín entero siempre van a ser bosones, y que las partículas de espín semientero siempre van a ser fermiones. Es un resultado, no una definición: el **teorema de la estadística del espín**. A pesar de la absoluta importancia de este teorema para el éxito de la teoría cuántica de campos, su demostración es un poco delicada desde el punto de vista matemático. Casi ningún libro de texto moderno explica el argumento, sino que se remite al texto clásico de 1964 *PCT, Spin and Statistics, and All That*, de Ray Streater y Arthur Wightman. Tampoco vamos a demostrarlo rigurosamente aquí, pero podemos ver por qué tal resultado podría ser plausible.

Para llegar a ese punto, pensemos un poco más en lo que significa realmente «espín». Para un objeto clásico, el espín es simplemente una velocidad de rotación alrededor de un eje. Las leyes de la naturaleza presentan una simetría en el sentido de que podemos girar nuestro marco de referencia y las leyes subyacentes no cambian. El teorema de Noether implica además la existencia de una cantidad conservada, que es el momento angular del objeto. A diferencia de los objetos clásicos, las partículas cuánticas pueden tener un espín intrínseco que nunca cambia su magnitud total, solamente su orientación en el espacio. El electrón tiene espín ½, el fotón tiene espín 1, y así sucesivamente. (Estamos utilizando unidades donde $\hbar = 1$, por lo que el momento angular de un electrón es en realidad $\hbar / 2$).

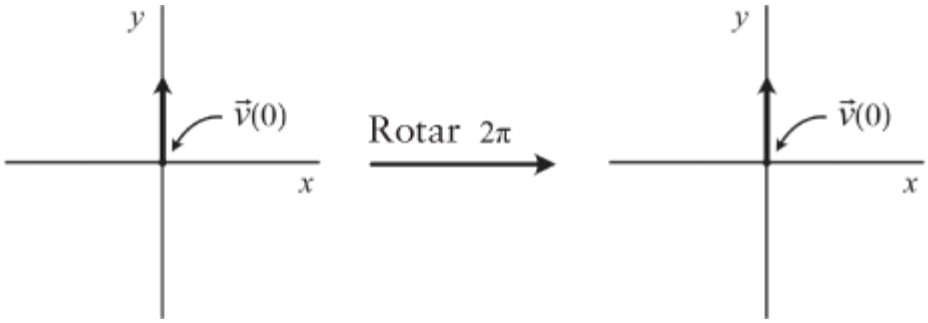
Ahora queremos ampliar esto a los campos cuánticos. Empezaremos preguntándonos qué le ocurre a un campo cuando realizamos una rotación en el espacio. De hecho, centrémonos en lo que le ocurre al campo en un único punto cuando realizamos una rotación alrededor de ese punto.

Es probable que su respuesta intuitiva sea «nada en absoluto», ya que una rotación alrededor de un punto concreto deja ese punto inmóvil. Y eso es cierto si pensamos en campos escalares, que son simplemente valores numéricos en cada punto. Ese valor no cambia,

por lo que el campo en sí no cambia, independientemente del ángulo de rotación.

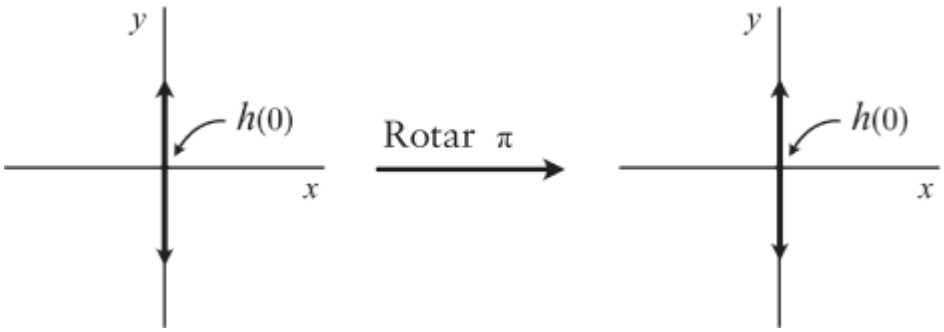


Cuando pasamos de los campos escalares, las cosas se complican. Pensemos en un campo vectorial $\vec{v}$. En cada punto tiene una dirección en el espacio y una magnitud. Cuando rotamos alrededor de ese punto, las componentes del vector cambian, a diferencia del valor de un simple escalar. Pero hay un caso especial, que es la rotación de exactamente 2π radianes (360 grados). En ese caso, el vector vuelve exactamente al punto en el que empezamos.



¿Podría haber otros ejemplos en los que rotar una cierta cantidad nos devuelva al punto de partida? Pues sí, ciertamente puede haberlos, y ciertos tipos de campos tensoriales cumplen los requisitos, incluidos los que describen las ondas gravitatorias (o gravitones, en el lenguaje cuántico). Un campo de este tipo tiene la propiedad que vuelve a su estado original tras una rotación de solamente π radianes, que es la mitad del círculo. Si esto parece difícil de visualizar, piense en el campo tensorial h como un segmento de línea con flechas en ambos

extremos, centrado en el punto que estamos considerando. Entonces queda claro que una rotación de π devuelve el campo en ese punto a su valor original.



La gran revelación es que estas propiedades de los campos bajo rotaciones están directamente ligadas al espín de la partícula cuántica asociada. El escalar, espín 0, es invariante bajo cualquier rotación. El vector, espín 1, es invariante bajo una rotación de 2 π. Y el gravitón, espín 2, es invariante bajo una rotación de π. La regla general es:

Una partícula de espín s es invariante bajo una rotación de $2\pi/s$ radianes.

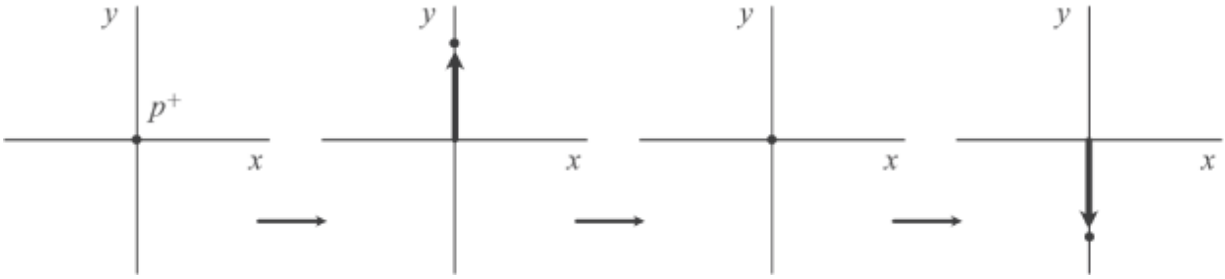
Lo interpretamos como«cualquier rotación»cuando $s = 0$.

CAMPOS CLÁSICOS Y ESPÍN

A veces es útil comprobar la realidad en estos dominios abstractos de la teoría cuántica de campos. Hagamos un breve paréntesis para ver cómo el comportamiento del espín de los fotones y gravitones se refleja en la dinámica de las ondas electromagnéticas y gravitatorias clásicas.

Primero, el electromagnetismo. Una onda viajera está formada por un campo eléctrico y un campo magnético, perpendiculares entre sí y que oscilan hacia arriba y hacia abajo al avanzar. Centrémonos en el campo eléctrico y su efecto sobre un protón que lo atraviesa. En la posición del protón, la onda electromagnética que pasa parece un

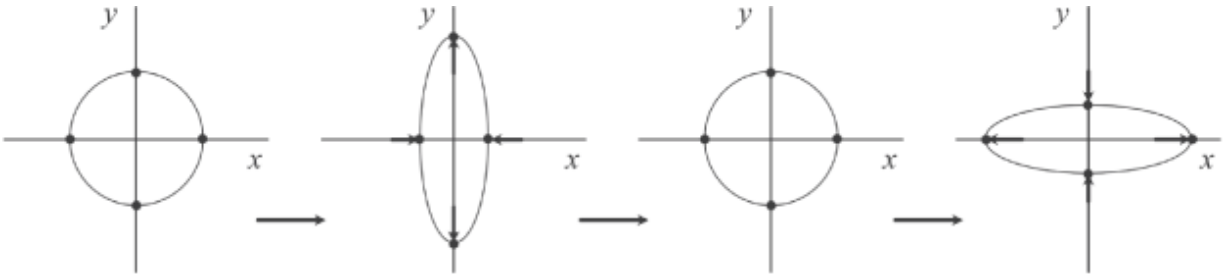
campo eléctrico que oscila hacia arriba, luego hacia abajo y vuelve hacia arriba. El protón es acelerado en la dirección de ese campo, por lo que será empujado hacia arriba y luego hacia abajo en un patrón repetitivo.



Si nos concentramos en la dirección de la fuerza eléctrica sobre el protón, vemos que en cada instante de tiempo, el patrón es invariante bajo una rotación 2π. Eso es justo lo que anticipamos para una partícula de espín 1, como lo es el fotón.

Ahora contrastemos eso con una onda gravitatoria. En este caso no podemos concentrarnos simplemente en una sola partícula. Cuando hablamos del principio de equivalencia en *Espacio, tiempo y movimiento*, vimos que un campo gravitatorio es esencialmente indetectable en un único punto del espacio, y eso es tan cierto para una onda gravitatoria como para el campo gravitatorio de una estrella o un planeta. Dada una sola partícula, siempre podemos elegir coordenadas que permanezcan centradas en la partícula, de modo que no parezca que se «acelere» en absoluto. En su lugar, tenemos que observar una colección de partículas cercanas y considerar sus movimientos relativos, que podemos tratar como fuerzas de marea de la onda gravitatoria que pasa.

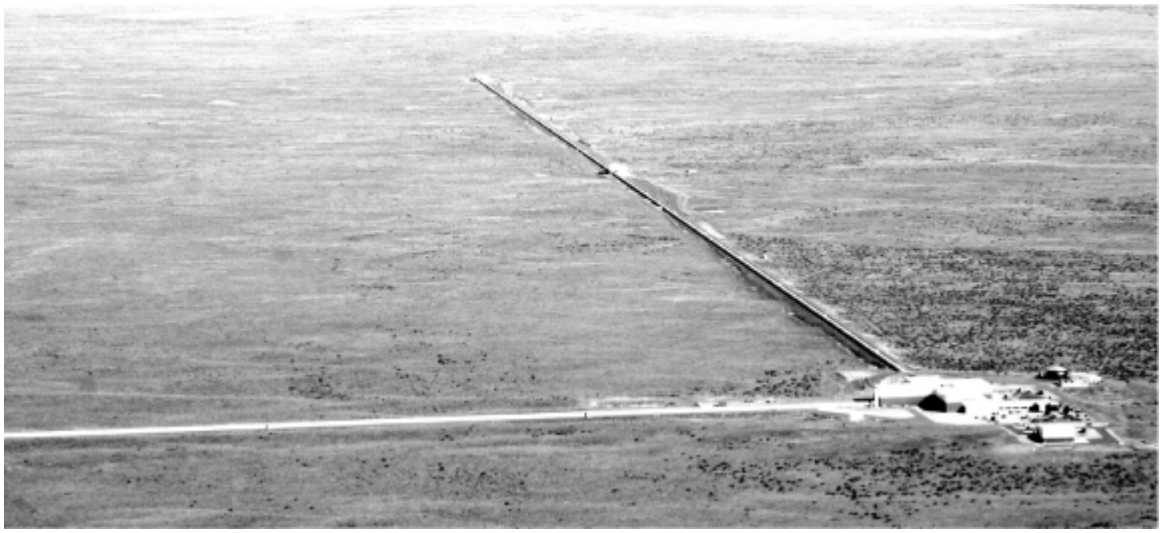
Consideremos un anillo de partículas flotando en el espacio al paso de una onda gravitatoria. Lo que ocurre es que la onda primero estira la configuración en una dirección mientras la comprime en la dirección perpendicular, luego se vuelve a relajar y estira/comprime en el sentido opuesto, continuando en una secuencia oscilante.



Así que cuando observamos el patrón de fuerzas de marea en un momento dado, nos damos cuenta de que es invariante bajo una rotación de π radianes, en lugar de 2π como en el caso electromagnético. Por eso el gravitón es una partícula de espín 2.

Hay que destacar dos cosas sobre esta sencilla demostración. En primer lugar, es posible que perciba cierta familiaridad entre la imagen de las fuerzas debidas al paso de una onda gravitatoria y una posible oscilación de un bucle circular de cuerda. No se trata de una coincidencia. La razón por la cual la teoría de cuerdas, que originalmente fue concebida como una aproximación a las interacciones fuertes, termina siempre como teoría de gravitación, es que este modo concreto de vibración de una cuerda cerrada actúa como una partícula sin masa de espín 2, la que interpretamos como el gravitón. (Para ser un gravitón hace falta algo más que tener el espín correcto: lo más importante es acoplarse a todas las formas de energía-momento, para satisfacer el principio de equivalencia, pero la teoría de cuerdas predice también esas propiedades).

Otra cosa es pensar en los observatorios de ondas gravitatorias, como la instalación LIGO de Hanford (Washington). Consta de dos brazos de cuatro kilómetros de largo que se extienden en ángulo recto desde una instalación central. Esos brazos son tubos de vacío por los que se disparan láseres que rebotan en espejos situados en los otros extremos y regresan, y en los cuales pueden medirse con una precisión fantástica los cambios en la distancia recorrida. Ahora ya sabe por qué se necesitan dos brazos y por qué es tan eficaz colocarlos en ángulo recto: porque una onda gravitatoria que pasa estirará las cosas en una dirección y las comprimirá en la perpendicular.



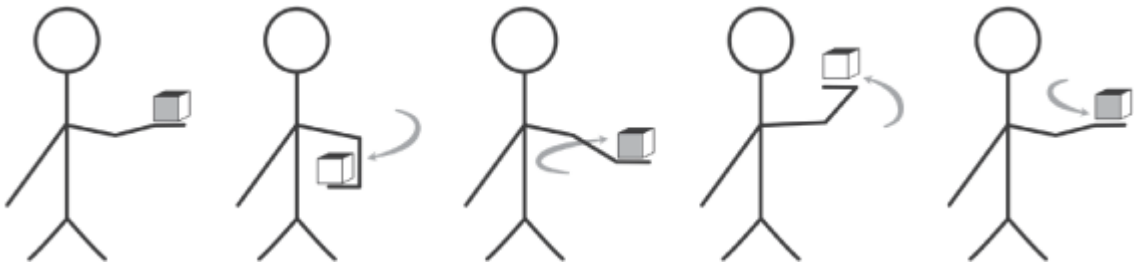
ESPÍN ½

Volvamos a la teoría cuántica de campos. Dimos una regla según la cual una partícula de espín *s* es invariante bajo una rotación de $2\pi/s$ radianes. El electrón es una partícula de espín ½, por lo que debería ser invariante bajo una rotación de 4π, que equivale a dar dos vueltas alrededor de un círculo. ¿Qué clase de monstruosidad no es invariante cuando la rotamos 2π radianes, pero es invariante cuando la rotamos 4π radianes?

No es tan difícil de imaginar algo así como podría pensarse en un principio. Pensemos en una banda de Möbius, una simple banda en la que dos extremos están conectados con un medio giro entre ellos. Si se imaginara caminando a lo largo de una banda de Möbius, después de dar una vuelta se encontraría en el lado opuesto al que había empezado. Únicamente después de dar dos vueltas se vuelve a la orientación original.

Otro ejemplo que se acerca más a la noción de «rotación» es sostener un cubo (o una taza de café, o lo que sea) en la mano y girarlo torciendo el brazo sin inclinar el objeto a lo largo de ningún otro eje. Puede mantener la orientación vertical del cubo moviéndolo, por ejemplo, por debajo del brazo y volviendo después. Su brazo se

torcerá un poco —por favor, inténtelo en casa, pero no se haga daño— y debería quedar claro que no hay forma de deshacer el giro del brazo sin mover el cubo. Pero podemos seguir rotando el cubo en la misma dirección —sin regresar por el mismo camino, eso sería hacer trampas —, esta vez por encima del brazo en lugar de por debajo. Milagrosamente, el cubo y el brazo volverán a su configuración original, después de haber girado 4π radianes. El cubo, considerado aisladamente, es invariante bajo rotaciones de 2π. Pero la relación entre el cubo y el mundo exterior, aquí representada por su brazo, solamente es invariante bajo rotaciones 4π. (Este movimiento forma parte de la danza balinesa de la vela, pero se popularizó entre los físicos en una conferencia de Richard Feynman).



El tema técnico que nos ocupa es la **teoría de la representación** del grupo de rotación SO(3). El grupo en sí es un conjunto abstracto de transformaciones; lo que realmente le ocurre a un objeto bajo tal transformación es la «representación» bajo la cual se transforma. Los escalares no cambian en absoluto; los vectores vuelven a su configuración tras una rotación de 2π; la representación particular bajo la que se transforman los gravitones vuelve a su configuración tras una rotación de π.

Los electrones y otras partículas de espín ½ (que incluyen todos los fermiones fundamentales descubiertos hasta la fecha) se transforman bajo una representación que vuelve a su configuración original tras una rotación de 4π. Estos objetos se denominan **espinores**.

Si el campo del electrón ψ_e no se ve afectado por una *rotación de* 4π, ¿qué le ocurre bajo una rotación de 2π, solamente una vez

alrededor del círculo? Probablemente pueda adivinarlo: obtiene un signo menos.

$$R\ (2\pi)\cdot\psi_e(x) = -\psi_e(x);\ \ R\ (4\pi)\cdot\psi_e(x) = \psi_e(x). \quad (11.5)$$

Eso tiene sentido, ya que hacer una rotación de 2π dos veces debería equivaler a hacer una rotación de 4π una vez. También es un poco provocador. Cuando hablamos de los fermiones, su característica definitoria era obtener un signo menos cuando se intercambiaban dos partículas idénticas. Ahora, con las partículas de espín ½, vemos que obtienen un signo menos bajo una rotación de 2π. Procure recordar esto.

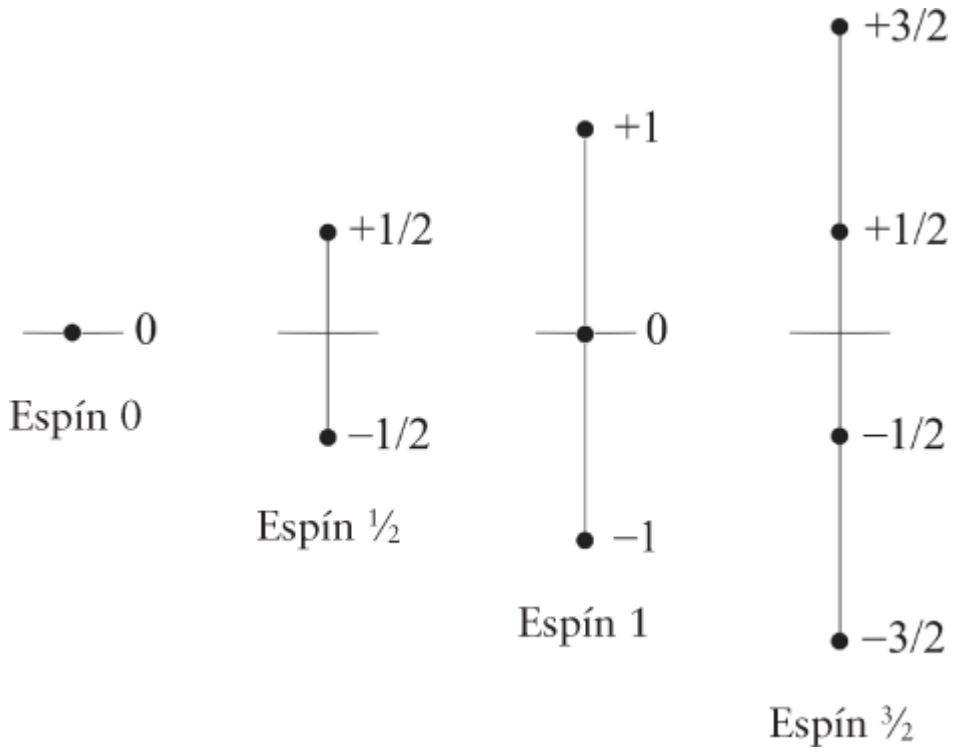
MEDICIÓN DEL ESPÍN

Cuando hablamos por primera vez del espín en el capítulo 2, mencionamos el sorprendente resultado del experimento de Stern-Gerlach, que enviaba un haz de partículas de espín ½ a través de un campo magnético colimado. Las partículas que tienen giro de espín actúan como pequeños imanes por sí mismas, de modo que al atravesar un campo magnético externo se desvían, dependiendo de cómo esté orientado su espín. Pero no obtenemos un abanico de partículas desviadas en ángulos diversos: lo que vemos son partículas desviadas hacia arriba en un ángulo fijo o hacia abajo en un ángulo fijo. Hemos medido el espín de la partícula a lo largo del eje elegido, y los dos posibles resultados de la medición son espín hacia arriba o espín hacia abajo. Las medidas de espín dan resultados cuantizados para la cantidad de momento angular a lo largo del eje de medida.

Pero no todas las partículas tienen espín ½. ¿Qué ocurre en circunstancias más generales?

Sigue habiendo resultados cuantizados, pero las posibilidades dependerán del espín de la partícula. La regla es que los posibles

resultados están separados por una unidad de espín (es decir, por $\hbar$), y que se extienden de menos el espín a más el espín. Esto significa que para el espín ½ los resultados posibles son –½ o +½ para la componente del momento angular a lo largo del eje de medición. Para una partícula de espín 0, no hay desviación. Pero para una partícula de espín 1, hay tres resultados posibles: –1, 0 y +1. Puede desviarse hacia abajo, hacia arriba o no desviarse en absoluto. Para una hipotética partícula de espín $\frac{3}{2}$ habría cuatro resultados posibles: $-\frac{3}{2}$, $-\frac{1}{2}$, $+\frac{1}{2}$ y $+\frac{3}{2}$. Para una partícula de espín s habrá $2s+1$ resultados posibles, que van de $-s$ a $+s$ en incrementos de 1.



Sin embargo, las partículas sin masa son un caso especial. Los $2s+1$ resultados posibles son los que podríamos medir en el marco de reposo de la partícula. Pero las partículas sin masa no tienen un marco de reposo, ya que siempre se mueven a la velocidad de la luz. En ese caso, solamente pueden presentar dos giros de espín diferentes: con su eje de espín apuntando en la dirección de su movimiento, o en dirección opuesta. Estos son los estados de «helicidad» de las

partículas sin masa de los que hablamos en el capítulo anterior. Así, mientras que un bosón masivo de espín 1 tiene tres estados de espín posibles, una partícula sin masa de espín 1 (como un fotón o un gluon) solamente tiene dos.

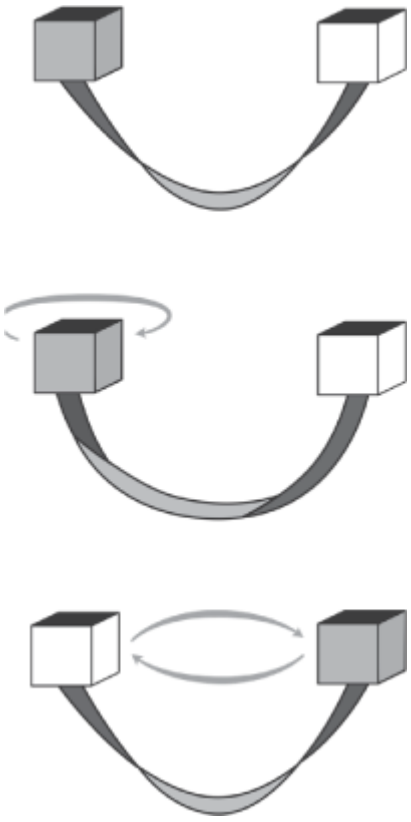
Esto explica un enigma persistente del capítulo anterior cuando hablamos del mecanismo de Higgs. Cuando se rompe espontáneamente una simetría de gauge, en lugar de obtener bosones de Goldstone sin masa, como ocurre cuando se rompe una simetría global, los bosones de gauge se vuelven masivos. Los supuestos bosones de Goldstone han sido «devorados» por los bosones de gauge. Pero no desaparecen del todo, ya que un bosón de gauge masivo de espín 1 tiene tres estados de espín posibles en lugar de solamente dos. El grado de libertad que habría supuesto un bosón independiente sin espín es ahora el componente de espín 0 del bosón de gauge masivo. Sigue habiendo el mismo número de cosas que pueden suceder en el mundo (variedades de campos vibratorios); solamente se organizan de forma diferente tras la ruptura espontánea de la simetría.

EL TEOREMA DE LA ESTADÍSTICA DE ESPÍN

Hemos visto que el intercambio de dos fermiones idénticos multiplica la función de onda por un signo menos, mientras que con la rotación de 2π de una sola partícula de espín ½ también se obtiene un signo menos. La esencia del teorema de la estadística de espín —todos los bosones tienen espín entero, los fermiones tienen espín semientero— es que estos dos signos menos son secretamente el mismo signo menos. Para la función de onda de varias partículas idénticas, intercambiar dos de ellas equivale a rotar solamente una. Puesto que los fermiones reciben un signo menos cuando se intercambian, también tiene que ser cierto que reciben un signo menos cuando se rotan 2π.

No vamos a demostrar el teorema, que se basa en detalles de la teoría cuántica relativista de campos, ni siquiera vamos a esbozar el razonamiento que hay detrás de la demostración. En su lugar, nos limitaremos a explicar cómo, en la idea del ejemplo de rotación de un cubo con el brazo que vimos, tiene mucho sentido que la rotación y el intercambio puedan ser equivalentes, aunque sean operaciones físicamente distintas.

Considere dos cubos, con una cinta que los conecta sin estar tensa. Se supone que los cubos son partículas idénticas, pero los hemos sombreado de forma diferente para que pueda ver cuándo se mueven. Partiendo de una configuración sin ninguna torsión en la cinta, vamos a rotar uno de los cubos un ángulo 2π, manteniendo fijo el otro cubo. Para concretar más, gírelo en el sentido de las agujas del reloj visto desde arriba. Ahora hay una torsión en la cinta: está en una configuración topológicamente diferente de la que teníamos al principio.



Ahora vamos a intercambiar los dos cubos entre sí, de nuevo en el sentido de las agujas del reloj, visto desde arriba. Mantenemos los cubos en una orientación constante a lo largo de la trayectoria, y a una altura fija, por lo que no se producen rotaciones adicionales. Pero está claro que la cinta ha vuelto a su configuración original. El intercambio de las dos partículas ha deshecho la torsión que introdujimos al rotar solamente una de ellas. Así es como se comportan los fermiones, al menos en un universo relativista con tres dimensiones espaciales.

MATERIA

El hecho de que los electrones sean fermiones explica por qué la materia es sólida. Como los electrones obedecen a la estadística de Fermi-Dirac, sabemos que no hay dos que puedan ocupar el mismo estado cuántico: el principio de exclusión de Pauli. Sin embargo, los físicos tardaron bastante tiempo en establecer cuidadosamente la conexión. Las ideas básicas fueron elaboradas por el físico matemático Elliott Lieb en la década de 1970.

Para ser sinceros, la conclusión no es obvia. El principio de exclusión no dice que dos partículas no puedan estar en el mismo *lugar*, sino que no pueden estar en el mismo *estado cuántico*. En el caso de dos átomos de hidrógeno, podríamos imaginar que los átomos se superpusieran de forma que sus electrones estuvieran casi en el mismo estado, aunque no del todo. Pero la mecánica cuántica no va a permitir que nos salgamos con la nuestra con ese truco. La exigencia real del principio de exclusión es que dos electrones cualesquiera tienen que estar en estados cuánticos completamente ortogonales. Si se intenta que dos de ellos coincidan casi en la misma configuración, uno o ambos serán empujados a un estado de mayor energía. Esto crea una repulsión entre los electrones conocida como **fuerza de intercambio o repulsión de Pauli**. Esta fuerza de repulsión es la responsable de que

los átomos ocupen espacio y, por tanto, del carácter sólido de la materia. (Para los bosones existe una fuerza atractiva análoga. Si los electrones hubieran sido bosones, el universo habría sido un lugar completamente distinto).

Es posible que dude al utilizar la palabra «fuerza», ya que es tradicional identificar solamente cuatro fuerzas fundamentales: la gravitación, el electromagnetismo y las fuerzas nucleares fuerte y débil. Pero esa identificación siempre fue casual y sugerente, no pretendía ser rigurosa. Como hemos dicho, en realidad todo son campos cuánticos y sus interacciones. La repulsión entre fermiones idénticos es solamente parte de la dinámica de un campo cuántico. Resulta útil pensar en ella como una «fuerza» cuando nos soltamos la melena y hablamos como si el mundo estuviera hecho de partículas clásicas, lo que es perfectamente aceptable en las circunstancias apropiadas.

En la mayoría de los materiales ordinarios, el principio de exclusión explica por qué los átomos ocupan espacio, pero los detalles de los átomos individuales y cómo se unen para formar moléculas también desempeñan un papel importante en las propiedades estructurales de cualquier sustancia dada. Por supuesto, también existen circunstancias extremas. Cuando una estrella típica agota su combustible nuclear disponible y ya no se sustenta con el calor generado por la fusión, se desprende de sus capas exteriores y el núcleo remanente se estabiliza en un estado conocido como **enana blanca**. La materia de una enana blanca es extremadamente densa, unas 200.000 veces más que la Tierra. La estrella se sostiene gracias a la **presión de degeneración** causada por el hecho de que todos sus electrones están empaquetados lo más apretadamente posible, dado que son fermiones. Son los electrones los que cuentan, no los núcleos atómicos, ya que los primeros son más ligeros y, por tanto, tienen mayores longitudes de onda Compton.

Si la masa de la estrella degenerada supera el **límite de Chandrasekhar**, protones y electrones se combinarán para formar neutrones (emitiendo neutrinos en el proceso). Los neutrones también

son fermiones, pero masivos y, por tanto, tendrán una longitud de onda Compton menor, así que lo que nos queda es una **estrella de neutrones**, aún más pequeña y más densa. Una enana blanca con la misma masa que el Sol tendrá aproximadamente el tamaño de la Tierra, mientras que una estrella de neutrones de la misma masa será solamente un poco más grande que la ciudad de Baltimore.

A masas aún mayores, ni siquiera la presión de los neutrones puede resistir la atracción de la gravedad. No se conocen configuraciones de materia ordinaria que tengan una densidad superior a esa, por lo que la estrella colapsará formando un agujero negro. Eso no viola el principio de exclusión: algunas partículas se verán forzadas a pasar a estados de mayor energía durante el colapso, hasta que finalmente alcancen la singularidad y la teoría cuántica de campos ya no sea capaz de describir lo que ocurre.

DOCE
ÁTOMOS

La palabra «átomo» nos llega de la Antigua Grecia. En el siglo V a.C., Demócrito (y su maestro, Leucipo, cuyos escritos se han perdido) sostenían que la materia no puede subdividirse indefinidamente y, por tanto, debe estar formada por algún tipo de partículas básicas indivisibles. Este concepto fue adoptado por los químicos del siglo XVIII, que consideraron que la mejor manera de explicar las proporciones de las sustancias en determinadas reacciones químicas era plantear la existencia de bloques de construcción elementales de los que están hechos todos los elementos. Tenían razón en lo que respecta a los elementos químicos, pero los físicos pronto comprendieron que los átomos pueden descomponerse en partículas aún más pequeñas: protones, neutrones y electrones, y que los protones y nucleones están formados por quarks *up*, quarks *down* y gluones. Ahora consideramos que las partículas elementales, que son a su vez excitaciones de campos cuánticos, son los componentes básicos de la materia.

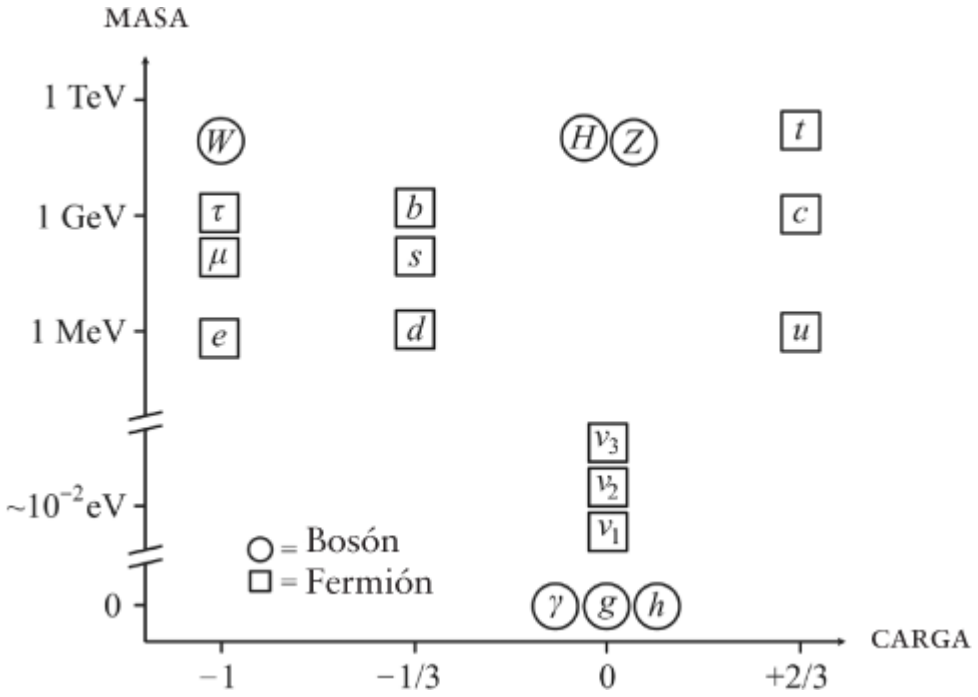
No obstante, es evidente, y sin duda importante para la organización de las cosas en el universo, que el mundo de nuestra experiencia inmediata presenta ciertas combinaciones de partículas elementales y no otras. ¿Por qué son tan comunes los electrones y los quarks *up*, y por qué son tan escasos los muones y los quarks *top*? Y dados los ingredientes básicos de los electrones y los quarks *up*/*down*, ¿por qué se ensamblan formando las configuraciones concretas que observamos?

LA TEORÍA DEL NÚCLEO ATÓMICO

Seamos primero un poco más sistemáticos sobre lo que son realmente los campos cuánticos conocidos y sus partículas asociadas. Los físicos suelen hablar del «Modelo Estándar», que no incluye la gravedad, ya que el Modelo Estándar es una teoría cuántica de campos renormalizable, y la versión cuántica de la relatividad general no sería renormalizable. Habiendo llegado a este punto en el libro, sabemos que lo que importa no es la renormalizabilidad, sino la existencia de una teoría de campo efectivo para partículas y procesos por debajo de algún corte ultravioleta. Y si eso es lo que nos interesa, no hay ningún problema en incluir la gravedad. Podemos tomar la métrica espacio-temporal $g_{\mu\nu}$ y escribirla como la suma de una métrica de fondo más una pequeña perturbación, $g_{\mu\nu} = g^0{}_{\mu\nu} + h_{\mu\nu}$, $g^0{}_{\mu\nu}$ es la métrica del espacio-tiempo plano de Minkowski. Entonces la perturbación h_{mn} puede tratarse como un campo ordinario que se propaga en el espacio-tiempo plano siempre que nos mantengamos alejados de los agujeros negros y otros fenómenos gravitatorios extremos. Su partícula asociada es, por supuesto, el gravitón. El modelo resultante ha sido bautizado por Frank Wilczek como **teoría del núcleo atómico**.

Muchos físicos han propuesto ingeniosos diseños geométricos para ayudarnos a visualizar todas las partículas de la teoría del núcleo atómico. La realidad es que es un lío. No pasa nada, nadie espera que sea la bella y definitiva Teoría del Todo, si es que existe tal cosa. La teoría del núcleo atómico no solamente no describe los agujeros negros ni el Big Bang, sino que tampoco da cabida a la materia oscura, que constituye la mayor parte de la materia del universo, y hay varios ajustes finos inexplicables y problemas de naturalidad que siguen causándole molestias. Su principal virtud es que se ajusta a los datos, lo que no deja de ser una virtud considerable. Así que vamos a aceptar el lío y a considerar todas las partículas de la teoría del núcleo atómico por turnos.

La figura muestra estas partículas, destacando tres propiedades básicas: la carga eléctrica, la masa y si son fermiones o bosones. Todos los bosones tienen espín 1, excepto el bosón de Higgs *H*, que tiene espín 0, y el gravitón *h*, que tiene espín 2. Todos los fermiones son de espín ½. Únicamente estamos representando partículas y no sus antipartículas, lo que implica algunas decisiones arbitrarias; por ejemplo, hemos mostrado el bosón *W* con carga negativa, pero no su homólogo con carga positiva.



Es una colección desordenada pero, en última instancia, manejable. Entre los bosones tenemos el bosón de Higgs *H* (masa 125 GeV); los bosones débiles *W* (80 GeV) y *Z* (91 GeV); y el fotón sin masa *g*, el gravitón *h* y ocho gluones *g*. Todos ellos son eléctricamente neutros excepto el *W*, con carga – 1. Existen dos tipos de fermiones, a saber, los quarks de interacción fuerte y los leptones, sin interacción fuerte, y dentro de cada categoría hay dos subconjuntos que difieren en una unidad de carga. Las partículas diferentes dentro del mismo subconjunto se denominan con distintos «sabores». Así pues, tenemos los quarks *down d* (4,7 MeV), *strange s* (96 MeV) y *bottom b* (4,2

GeV), con carga – 1/3, y los quarks *up* u (2,2 MeV), *charm* c (1,3 GeV) y *top* t (173 GeV), todos ellos con carga +2/3. Los leptones cargados incluyen el electrón e (0,51 MeV); el muon μ (106 MeV); y el tau t (1,8 GeV), todos ellos con carga – 1. Luego hay tres neutrinos ν, todos ellos, por supuesto, neutros. Normalmente se oye hablar de los tres tipos de neutrinos asociados a los leptones cargados: hay neutrinos electrónicos, neutrinos muónicos y neutrinos tau, que indican qué tipo de neutrinos se crean en las interacciones débiles. Pero los neutrinos son bestias escurridizas, y estos tres tipos son en realidad combinaciones de sabores más básicos ν_1, ν_2 y ν_3. Cada uno de estos últimos tiene una masa definida, que sabemos que tiene que ser inferior a 1 eV, pero los experimentos actuales no han determinado valores precisos para las masas de los neutrinos.

Por último, tenemos la regla de que las partículas que interaccionan fuertemente se agrupan en combinaciones incoloras, los hadrones. Los dos tipos principales son los mesones, formados por un quark y un antiquark, y los bariones, formados por tres quarks. Se ha especulado con la posibilidad de que existan estados ligados formados exclusivamente por gluones, denominados **glubolas** (yo no tengo la culpa del nombre), pero si existen, van a ser de duración muy efímera.

LO QUE QUEDA

Nuestro universo está en expansión, por lo que hace mucho tiempo era mucho más denso y, por lo tanto, más caliente. Como regla general, si la temperatura es superior a la masa de una determinada partícula, esta se creará y aniquilará continuamente como resultado de interacciones térmicas aleatorias. (La temperatura y la energía tienen las mismas unidades, y ambas son equivalentes a la masa cuando fijamos $c = 1$, por lo que está bien comparar la masa de una partícula con la temperatura de lo que la rodea).

Por tanto, el universo primitivo estaba ocupado por todas las partículas conocidas en un plasma de alta temperatura, que podemos considerar como un gas que se ha calentado tanto que los átomos no pueden sobrevivir, dejando solamente partículas individuales libres. A medida que el plasma se expande y se enfría, la temperatura desciende por debajo de las masas de los distintos tipos de partículas. En ese momento, las partículas pueden aniquilarse, pero ya no hay energía suficiente en una colisión típica para crearlas, y desaparecen de la escena. No vamos a seguir ese proceso en detalle, sino que vamos a considerar lo que les ocurre a las partículas individuales cuando se las deja evolucionar por sí solas.

Tenemos las tres directrices siguientes:

1. Las partículas más pesadas se desintegran en otras más ligeras. En la relatividad especial, la energía se conserva, pero la «masa» no se conserva por separado, es solamente una forma de energía. Una partícula pesada puede desintegrarse y convertirse en un conjunto de partículas con una masa total menor, siendo la energía remanente la energía cinética de las partículas resultantes. También puede haber absorción, en la que varias partículas se convierten en una sola, pero esto es mucho menos frecuente que la desintegración. Las reglas subyacentes parecen las mismas hacia delante o hacia atrás en el tiempo, por lo que esto debe ser consecuencia del aumento de la entropía. En igualdad de condiciones, una partícula tiene menos entropía que un conjunto de varias partículas.
2. Las cantidades conservadas no desaparecen. Para nosotros, las cantidades relevantes van a ser la carga eléctrica, el número de Lepton (el número de leptones menos el número de antileptones) y el número de bariones (un tercio del número de quarks, es decir, el número de quarks menos el número de antiquarks).

3. Algunas partículas interaccionan y se adhieren entre sí formando sistemas compuestos como átomos y núcleos, mientras que otras pasan en gran medida unas a través de otras. Gravitones y neutrinos están pasando a través de su cuerpo ahora mismo, pero la probabilidad de que interaccionen con usted es insignificante. (Las partículas de materia oscura también podrían estar pasando a través de usted, pero en las Ideas Fundamentales nos ceñimos sobre todo a las cosas que conocemos).

Combinando las dos primeras de estas reglas, podemos esperar encontrar una gran cantidad de electrones, que son las partículas más ligeras con carga eléctrica; neutrinos, que son las partículas más ligeras con número de Lepton; y protones, que son las partículas más ligeras con número de bariones. Y precisamente esos tipos de partículas son bastante comunes en nuestro universo. Pero los detalles importan, entre ellos, por qué sigue habiendo neutrones en el mundo.

Otro hecho cosmológico digno de mención es la asimetría materia/antimateria. Según las leyes de la física conocidas, no hay razón para que haya más átomos de materia que de antimateria, pero en el mundo real claramente los hay. Incluso sabemos que otras estrellas y galaxias están hechas de materia y no de antimateria, porque, de lo contrario, esperaríamos una radiación de alta energía procedente de las regiones intergalácticas que hay entre ellas, donde la materia y la antimateria entrarían en contacto y se aniquilarían de forma gradual, pero segura. No vamos a explicar esto aquí, ya que la explicación es desconocida, sino que simplemente lo damos por sentado. Un detalle importante es que no sabemos con certeza que haya más materia que antimateria, ya que podría haber un baño de exceso de antineutrinos a nuestro alrededor. Lo que sí sabemos es que hay más bariones que antibariones, ya que no hay partículas casi invisibles en las que esconder el exceso del número de antibariones. Por esta razón, los físicos conocen este enigma como la **asimetría**

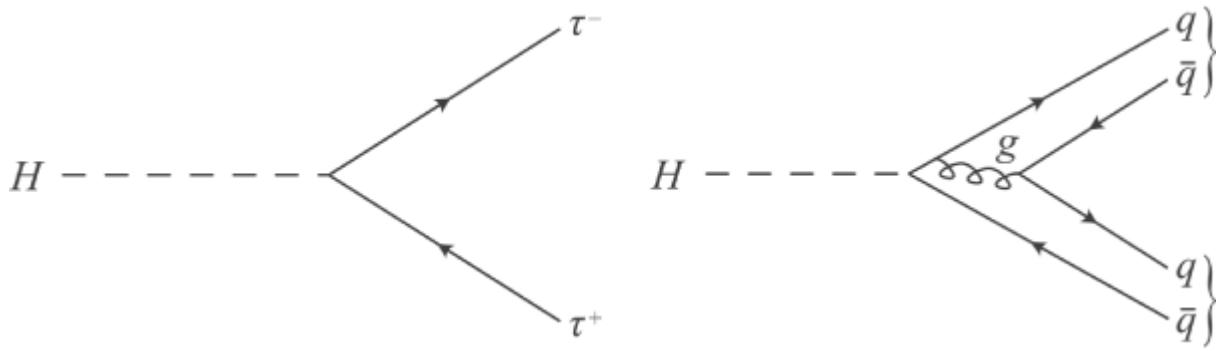
bariónica, y los mecanismos inventados para producirla se engloban bajo la rúbrica de la **bariogénesis**. La asimetría no es tan grande: por cada 10.000 millones de antibariones en el universo primitivo, había 10.000 millones y un bariones. Ese uno de más explica todos los bariones que vemos hoy en día.

MODOS DE DESINTEGRACIÓN

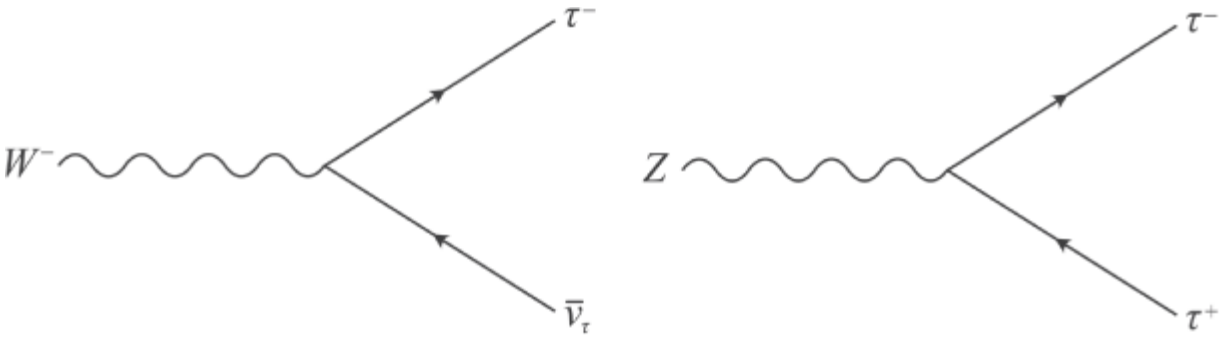
Veamos las distintas formas en que pueden desintegrarse las partículas. Como es nuestra costumbre, no vamos a ser completamente sistemáticos, sino que nos limitaremos a exponer las ideas centrales. Nuestra tarea se simplifica por el hecho de que los fotones, los gravitones y los neutrinos no tienen nada más ligero en lo que desintegrarse. Así que hay cuatro categorías de cosas de las que preocuparse: el bosón de Higgs, los bosones *W/Z*, los quarks y los leptones cargados.

Empecemos por el bosón de Higgs. Se trata de una partícula neutra, sin espín, por lo que su estrategia básica de desintegración será convertirse en un par partícula/antipartícula. Esto garantiza que todas las cantidades conservadas serán 0. Recuerde del capítulo 10 que los términos de masa para todas las demás partículas del Modelo Estándar se originan en su acoplamiento al Higgs, y en (10.8) vimos que las masas de los fermiones serían proporcionales a esos acoplamientos. Por tanto, el bosón de Higgs quiere desintegrarse en las partículas más pesadas que pueda: la constante de acoplamiento relevante es la misma cantidad que determina la masa de la partícula. (Se suprime la desintegración en *top* y *antitop* porque incluso un solo quark *top* es más pesado que el Higgs[1]). Así que al Higgs le gusta desintegrarse en un par tau/antitau, y aún más desintegrarse en un par *bottom/antibottom*. Puede que le preocupe esta última posibilidad, ya que se supone que los quarks están confinados en hadrones. Pues no es un problema: como las interacciones fuertes son fuertes, es bastante

fácil crear algunos quarks y gluones más por el camino, en un proceso llamado **hadronización**. Lo que observamos en realidad es la desintegración del bosón de Higgs en dos mesones o, de forma más realista, en una lluvia de mesones y bariones, además de otros fotones y leptones diversos que se crean como subproductos.



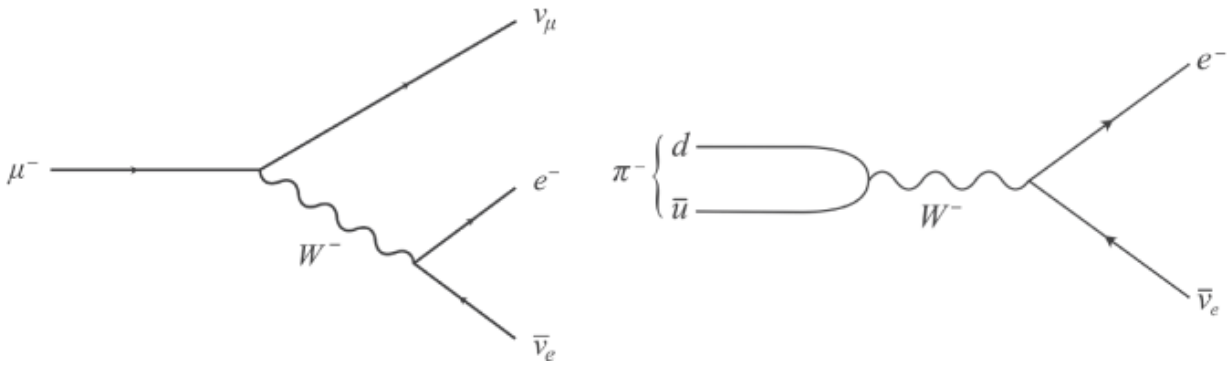
Los bosones *W* y *Z* van a desintegrarse de diferentes maneras, simplemente porque el *W* está cargado eléctricamente, y esa carga tiene que ir a alguna parte. Una desintegración típica de *W* va a ser en un fermión y un antifermión con cargas que difieren en 1, como un electrón y un antineutrino o un quark *down* y un quark *antiup* (que luego se hadronizarían). El *Z*, por su parte, se desintegra como el Higgs, en partículas y sus antipartículas, aunque con tasas de desintegración diferentes.



De hecho, los bosones *W* son las únicas partículas de la teoría del núcleo atómico que tienen un vértice fundamental que implica dos sabores distintos de fermiones. Los fotones, gravitones, gluones, el

bosón de Higgs y el bosón Z tienen vértices en los que se convierten en una partícula y su antipartícula exacta, pero no en una partícula y un sabor diferente de antipartícula. Tales procesos serían **corrientes neutras que cambian de sabor**, ya que todas las partículas en desintegración son eléctricamente neutras. Por lo que sabemos, nunca ocurren directamente en la naturaleza (es decir, no existe un vértice de diagrama de Feynman fundamental de ese tipo), pero pueden producirse a través de diagramas de bucle, aunque a una tasa inferior a la que cabría esperar de una interacción directa.

Por lo que respecta a los fermiones, sus modos de desintegración se producen generalmente mediante la emisión de un bosón W (virtual, para los fermiones más ligeros) y la conversión en un sabor diferente. A continuación, el bosón W se desintegra en otro fermión y un antifermión. Un muon, por ejemplo, se desintegra generalmente convirtiéndose en un neutrino y un W, y este último puede desintegrarse en un electrón y su antineutrino. La misma idea es válida para los quarks, incluso si están dentro de un hadrón: los neutrones se desintegran cuando uno de sus quarks *down* se convierte en un quark *up* y un W. Esta es incluso una forma común de desintegración de los mesones. Un mesón π^-, por ejemplo, es un estado ligado de un quark *down* y un *antiup*. Estos dos pueden aniquilarse produciendo un W^-, que luego se desintegra en un electrón y un antineutrino.



Como no hay corrientes neutras directas que cambien de sabor, los neutrinos más pesados no pueden desintegrarse directamente en

neutrinos más ligeros. Podemos imaginar diagramas de bucle que resolvieran el problema, pero las tasas de desintegración asociadas serían extremadamente pequeñas y nunca se han observado de modo experimental.

A todos los efectos, podemos tratar a todos los neutrinos como partículas estables.

Con todas estas desintegraciones permitidas, dada cualquier colección inicial de partículas (con cierto desequilibrio de bariones sobre antibariones), esperaríamos que se relajara rápidamente a una combinación de solamente los siguientes tipos:

— Bosones de gauge no confinados sin masa, el fotón y el gravitón
— Tres tipos de neutrinos
— Electrones
— Quarks *up* y *down* confinados en protones, las partículas más ligeras con número de bariones distinto de cero

Esto se corresponde bastante bien con el universo que vemos, con una excepción importante: los neutrones. Hay neutrones a nuestro alrededor; casi la mitad de la masa de un ser humano típico procede de los neutrones de sus núcleos atómicos. Tenemos que pensar un poco a fondo para entender por qué no se desintegran sin más.

NEUTRONES Y NUCLEIDOS

Las partículas más pesadas tienden a desintegrarse en otras más ligeras, dadas las limitaciones de las leyes de conservación. La masa de un neutrón es de 939,6 MeV. Puede, y así lo hace, desintegrarse en una combinación de un protón (938,3 MeV), que acapara el número de bariones; un electrón (0,51 MeV), que compensa la carga eléctrica del protón, y un antineutrino (<1 eV), que compensa el número de Lepton del electrón. Las cantidades conservadas permanecen constantes

durante la transformación, y la masa combinada de protón + electrón + neutrino es menor que la de un solo neutrón.

Pero, hay que reconocerlo, por muy poco. El neutrón no es mucho más pesado que la troika protón/electrón/antineutrino, como puede comprobarse considerando la diferencia fraccional de masa:

$$\frac{m_n - \left(m_p + m_e + m_{\bar{\nu}}\right)}{m_n} \approx 0{,}001. \qquad (12.1)$$

Intuitivamente, no hay mucho «espacio» para la desintegración de un neutrón: las energías cinéticas de las partículas creadas tienen que ser bastante pequeñas. Las matemáticas de la tasa de desintegración lo reflejan, y los neutrones acaban viviendo mucho tiempo (unos quince minutos) en comparación con otras partículas inestables, que suelen desintegrarse en minúsculas fracciones de segundo.

Si los parámetros fundamentales de la física de partículas hubieran sido solamente un poco diferentes, los protones podrían haber sido más pesados que los neutrones y se desintegrarían en ellos, en vez de al revés. Eso sería un universo muy aburrido. Nuestro universo es gratamente complejo porque hay muchos tipos de núcleos estables, que están positivamente cargados y pueden capturar electrones, lo que da lugar a abundantes formas de química. Si todos los núcleos fueran neutrones, serían eléctricamente neutros e incapaces de capturar electrones. No habría átomos, ni química, ni vida.

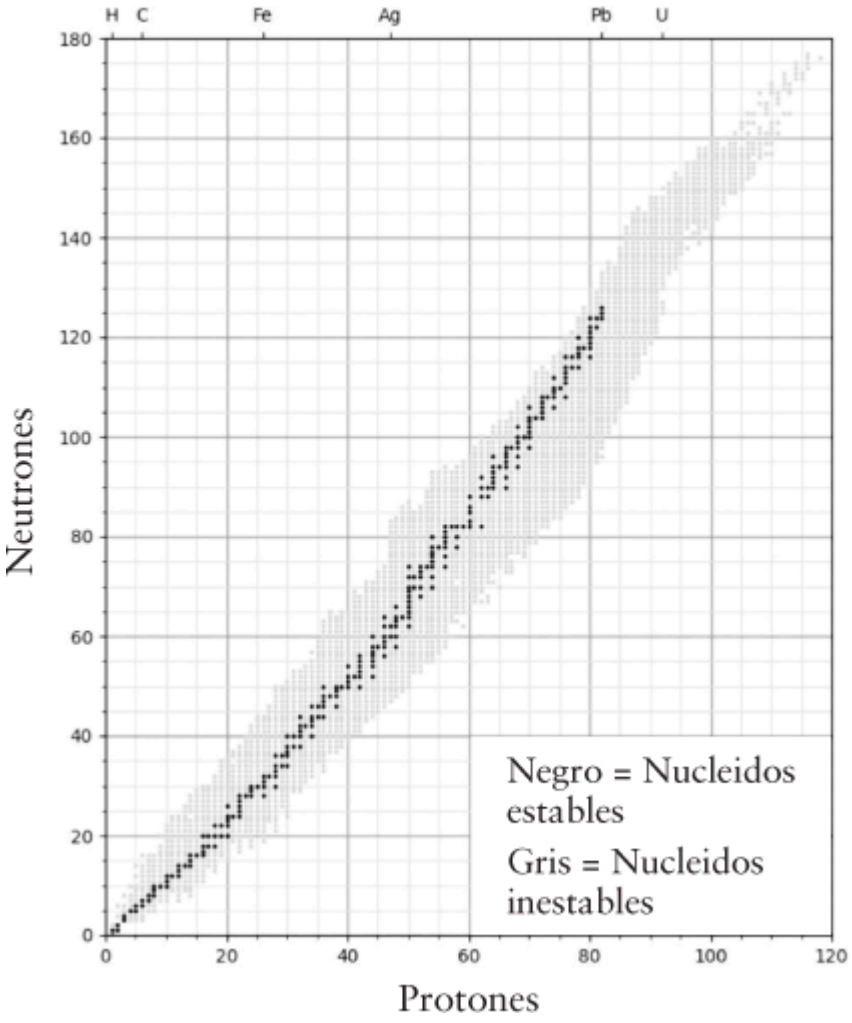
Pero ¿por qué existen los núcleos? ¿Por qué los neutrones de su interior no se desintegran en pocos minutos, en lugar de durar prácticamente para siempre?

En el capítulo 7 hablamos de la energía de enlace de un electrón con un protón: la masa de un átomo de hidrógeno enlazado es menor que la suma de las masas del electrón y del protón, porque hay una energía neta negativa en su atracción electromagnética. Si esto le

parece raro, es solamente una forma de decir que tendríamos que inyectar energía para romper el átomo.

La misma idea se aplica a los neutrones y protones de los núcleos. La fuerza de interacción fuerte es atractiva, y a los neutrones y protones les gusta estar bien pegados. Esto significa que la energía de un núcleo puede ser menor que la suma de los protones y neutrones que lo componen. Por ejemplo, el núcleo compuesto más simple es el **deuterón** (núcleo de hidrógeno pesado), que es solamente un protón y un neutrón. Tiene una masa de 1875,6 MeV, que puede considerarse como la masa del protón más la masa del neutrón menos 2,2 MeV de energía de enlace. Si el neutrón se convirtiera en un protón, un electrón y un antineutrino y las partículas siguieran caminos separados, la masa total tendría que ascender a 1877,1 MeV. No es posible desintegrarse en un conjunto de partículas más pesadas. Así que la energía de enlace hace que el deuterón sea estable, aunque un neutrón libre no lo sea.

Un tipo particular de núcleo —especificado por su número de protones y su número de neutrones— se denomina **nucleido**. La fuerza de interacción fuerte quiere unir a los nucleones, pero los protones tienen carga eléctrica positiva, lo que provoca una repulsión mutua. En consecuencia, el tamaño de un nucleido está limitado. Existen unos 250 nucleidos estables de forma efectiva —ya sean perfectamente estables o con una vida superior a la edad del universo— y muchos otros inestables.



EL MUNDO COTIDIANO

Una vez que todo alcanza el equilibrio, estamos en condiciones de describir los ingredientes básicos de los que estamos hechos usted y yo y todo lo que vemos directamente a nuestro alrededor. Olvídese de los neutrinos, que están ahí fuera en el universo, pero no afectan a nuestra vida cotidiana. Para el caso, olvídese también de las fuerzas nucleares fuerte y débil: la fuerza fuerte mantiene unidos los núcleos, pero una vez que sabemos que lo hace, la fuerza fuerte no tiene mucho que hacer en el mundo cotidiano, a menos que pasemos la vida trabajando en una central nuclear. (Las fuerzas nucleares desempeñan un papel crucial en el proceso de fusión que da energía al sol, así que en ese

sentido son importantes para nuestra existencia, pero no hace falta que las tengamos en cuenta al describir cómo nos comportamos nosotros mismos).

Lo que nos queda son dos tipos de partículas que componen la «materia»:

— Electrones
— 250 nucleidos estables

Y también tenemos dos fuerzas de largo alcance que son relevantes para la física macroscópica:

— Electromagnetismo
— Gravedad

Eso es todo, básicamente. Usted y todos sus amigos y seres queridos, por no hablar de sus conocidos ocasionales y enemigos acérrimos, son combinaciones de estas partículas de materia que interaccionan entre sí y con estas fuerzas según las reglas de la teoría cuántica de campos. Cada libro que ha leído, cada comida que ha ingerido, cada cuadro que ha visto y cada canción que ha escuchado han surgido de la dinámica colectiva de estas piezas básicas. Es un logro intelectual asombroso que los seres humanos hayan descubierto todo esto. Aún queda mucho trabajo por hacer para demostrar cómo estos constituyentes microscópicos se unen para producir las maravillas del mundo macroscópico, pero la lista de ingredientes se conoce bien.

Estos ingredientes no desempeñan el mismo papel. Entre las fuerzas hay una distinción crucial: la gravedad es algo torpe, mientras que el electromagnetismo es inteligente. La gravedad tiene muchas sutilezas cuando empezamos a pensar en cosmología y agujeros negros, pero aquí en la Tierra es bastante sencilla: todo tira de todo lo demás. Y ya está. Por eso no hay esperanza realista de crear un dispositivo antigravitatorio o maquinaria de manipulación de la

gravedad: la gravedad solo actúa en una dirección. Esto contrasta drásticamente con el electromagnetismo, donde tenemos cargas tanto positivas como negativas. Este simple hecho nos permite anular los campos electromagnéticos, focalizarlos y, en general, manipularlos con una precisión milimétrica. El electromagnetismo puede empujar y tirar, lo que le confiere una variedad de efectos enormemente mayor que los que obtenemos con la gravedad.

Del mismo modo, los electrones y los núcleos desempeñan papeles muy diferentes. La lista de nucleidos proporciona variedad: tenemos una colección de elementos químicos que podemos enlazar en maravillosas combinaciones moleculares. Pero, dados esos ingredientes, son los electrones los que hacen el trabajo más interesante, simplemente porque son mucho más ligeros y, por tanto, más fáciles de manipular. Son los electrones que interaccionan electromagnéticamente los que son movidos para crear electricidad, los que son repartidos entre los átomos para crear química y, en general, los que impulsan los procesos de la tecnología y la vida.

ELEMENTOS

Las cargas opuestas se atraen, por lo que a los núcleos y electrones les gusta asociarse para formar átomos. Los electrones son fermiones sujetos al principio de exclusión de Pauli. Esto hace que los átomos sean tan complicados como interesantes.

Los átomos de hidrógeno solamente tienen un electrón, que se encontrará en equilibrio en un único estado fundamental si se le permite. El deuterio (un protón y un neutrón) es más pesado que el hidrógeno ordinario, pero cuenta como el mismo elemento porque tiene el mismo número de protones y, por tanto, es muy similar en sus características atómicas. El siguiente elemento más pesado que podemos imaginar construido a partir de nuestros ingredientes es el helio, con dos protones y dos neutrones en el núcleo, es decir, dos

electrones en un átomo neutro. Según el principio de exclusión, esos dos electrones no pueden ocupar el mismo estado cuántico. Pero hay otra variable al acecho: el espín del electrón, que puede ser hacia arriba o hacia abajo. Por tanto, los dos electrones de un átomo de helio pueden tener exactamente el mismo perfil espacial, siempre que tengan espines opuestos. Sin embargo, no podemos acomodar más electrones en la misma función de onda espacial, u **orbital** como les gusta decir a los químicos. Esta es la razón por la que el helio es un **gas noble**: cada átomo tiene una dotación completa de electrones en su orbital más externo (y en este caso, el único), por lo que no siente ninguna necesidad de ganar, perder o compartir electrones con otros átomos y, por tanto, permanece libre de compromisos moleculares.

Los dos elementos siguientes son el litio y el berilio, con tres y cuatro protones, respectivamente. En cada uno de ellos, hay dos electrones acurrucados en el núcleo, como en el helio, pero también más electrones al acecho en estados de mayor energía más allá de ese orbital inicial. Por tanto, cabría esperar que el litio, con un único electrón fuera del primer orbital lleno, tuviera unas propiedades químicas parecidas a las del hidrógeno, y eso es correcto[2]. También cabría esperar que el berilio fuera un gas noble como el helio, pero eso es del todo erróneo.

La razón es que solamente hay una configuración espacial para los orbitales de menor energía, pero hay múltiples configuraciones posibles para los orbitales de energía inmediatamente superior. De hecho, hay cuatro: una que es esféricamente simétrica, pero que oscila radialmente desde el origen, y tres que no son esféricamente simétricas, sino que se extienden a lo largo de uno de los tres ejes espaciales. Así que, después del helio, podemos colocar $2 \times 4 = 8$ electrones nuevos en estos cuatro orbitales antes de llenarlo todo totalmente. El siguiente gas noble es en realidad el neón, con diez protones.

Y así sucesivamente, con más complicaciones por el camino a medida que pasamos a elementos más pesados y más orbitales de

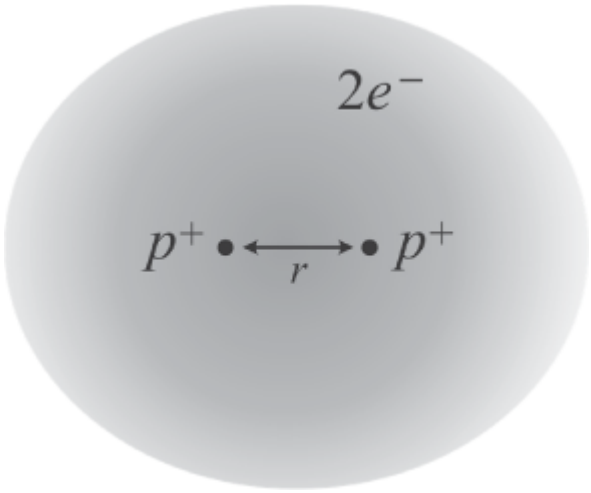
electrones permitidos. Las formas específicas en que los electrones pueden encontrarse alrededor de los núcleos sin dejar de ser compatibles con el principio de exclusión son las responsables de que los átomos sigan el patrón descrito por la **tabla periódica** de los elementos. Es todo muy intrincado y bello, y debería alegrarse de que esta serie de libros se concentre en la física fundamental y no en la química y los materiales y otros fenómenos colectivos, pues, de lo contrario, se extendería a un centenar de libros.

QUÍMICA

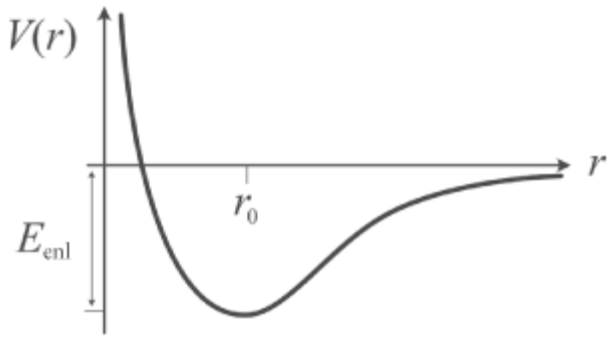
La diversión no acaba aquí. Desde el punto de vista de la física, cuando estructuramos una lista de nucleidos estables o los átomos en la tabla periódica, lo que realmente estamos haciendo es buscar configuraciones localmente estables de los campos de la teoría del núcleo atómico. Una configuración es «localmente estable» si tiene menos masa que cualquier otra configuración en la que podría transformarse fácilmente. Por regla general, eso significa que, o bien es realmente estable, o bien la única forma en que podría cambiar es pasando por otras configuraciones que tengan mayor masa. Eso puede ocurrir si se aplica un empujón adecuado. Una vela colocada sobre una mesa es localmente estable, pero si la enciende, empezará a arder, transformándose en un conjunto de productos de combustión con menor masa en conjunto y mayor entropía total.

Los átomos no son las únicas configuraciones localmente estables de nucleidos y electrones: también tenemos moléculas. Los químicos distinguen entre varios tipos de enlaces químicos, pero la idea básica es que podemos encontrar configuraciones de baja energía de electrones que orbitan alrededor de más de un núcleo. Consideremos dos átomos de hidrógeno, que pueden combinarse para formar una molécula de hidrógeno, H_2. Los electrones (con espines opuestos, para mantener contento a Pauli) pueden establecerse en un orbital

compartido alrededor de ambos protones. La energía resultante es unos 4,5 eV inferior a la de dos átomos de hidrógeno por separado, representando la diferencia la energía de enlace de la molécula.



En esta configuración, los protones estarán separados una distancia *r*. Si nos acercamos más, la repulsión electrostática entre los protones los separaría; si nos alejamos más, los electrones tirarían de ellos acercándolos. Podemos pensar en esto en términos de una energía potencial del sistema en función de la separación. La energía de enlace no es más que esta energía potencial, evaluada en su mínimo.



Empezamos este libro pensando en la emisión cuantizada de luz de los átomos, cuando un electrón salta de un orbital a otro. En las moléculas, de nuevo habrá niveles de energía excitados por encima del estado fundamental; como mencionamos en el capítulo 7, estos tienen

energías características del orden de 10^{-2} eV. Y, por supuesto, hay todo tipo de moléculas más complicadas que pueden formarse, interaccionando de diversas maneras en diferentes condiciones. Al final, no es más que electromagnetismo aplicado en un mundo con un par de cientos de nucleidos y el principio de exclusión de Pauli, pero un mundo así presenta, de hecho, una riqueza increíble.

LA FÍSICA DEL MUNDO COTIDIANO

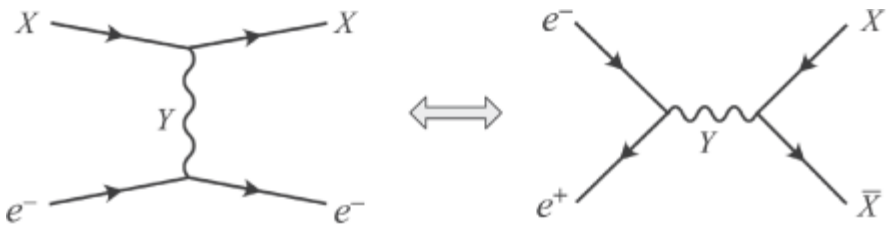
La teoría del núcleo atómico no es, obviamente, la teoría definitiva de la física. No describe la materia oscura ni las condiciones de los campos gravitatorios fuertes, y presenta varias coincidencias y refinamientos que sugieren una explicación más completa aún por llegar. La gente ha tenido algunas ideas sobre cómo podría ser esa teoría completa, pero hasta ahora no lo sabemos. Tampoco sabemos lo cerca que estamos de conocerla; puede que mañana se publique un artículo con la respuesta definitiva, o puede que dentro de mil años sigamos buscándola.

Pero hay un régimen en el que la teoría del núcleo atómico es totalmente exitosa. Si no tenemos campos gravitatorios fuertes, y solo consideramos los procesos que tienen lugar muy por debajo de algún corte ultravioleta prudentemente elegido (digamos, 10 GeV), la estructura básica de la teoría cuántica de campos nos asegura que la teoría del núcleo atómico es la historia completa. Sin duda, podemos imaginar otros campos que aún no hemos descubierto —la gran mayoría de los físicos apostaría a que hay muchos por ahí—, pero, o bien son tan masivos que nunca se producen en interacciones por debajo del corte, o bien se acoplan a la materia ordinaria tan débilmente que pueden ignorarse por completo[3].

Una de las razones por las que podemos hacer afirmaciones tan rotundas con tanta confianza es una característica de la TCC conocida como **simetría de cruce**. No se trata de una simetría de los campos

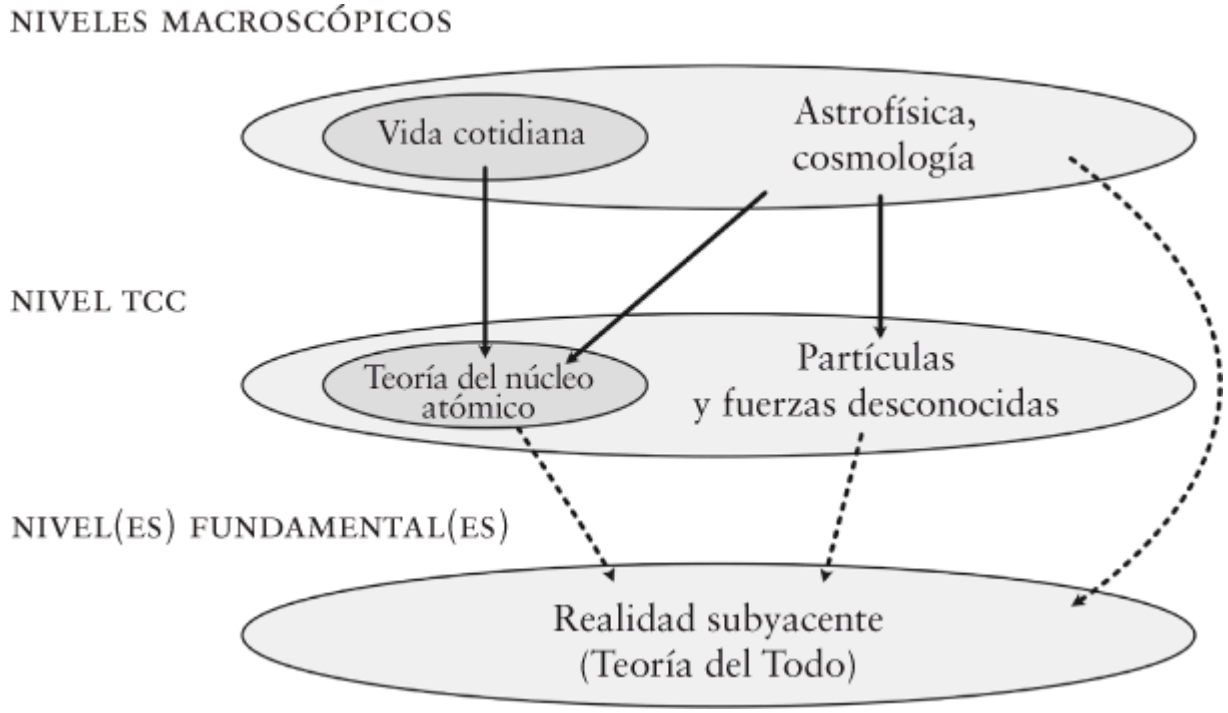
físicos, sino de los diagramas de Feynman que utilizamos para describir sus interacciones. A grandes rasgos, dice que podemos girar cualquier diagrama 90 grados —convirtiendo partículas en antipartículas cuando se invierte su dirección en el tiempo— y obtener otro diagrama con la misma amplitud.

Imaginemos que tenemos un nuevo fermión hipotético X, que suponemos que interacciona con los electrones mediante el intercambio de un nuevo bosón hipotético Y. La simetría de cruce dice que la existencia de una interacción $X + e^- \rightarrow X + e^-$ implica la existencia de una interacción relacionada $e^+ + e^- \rightarrow X + \overline{X}$. Así que si X interacciona notablemente con los electrones, podemos crear fácilmente X en el laboratorio haciendo colisionar electrones con positrones y observando lo que surge. Por fortuna, colisionar partículas y observar lo que surge es lo que más les gusta hacer a los físicos de partículas. Tenemos muy buenos datos sobre lo que se produce en estos procesos, y no se ha encontrado ninguna partícula misteriosa. Cualquier partícula que no esté ya en la teoría del núcleo atómico tendría que tener un acoplamiento realmente muy débil.



La notable consecuencia de esta línea de razonamiento es que tenemos excelentes razones para creer que las leyes de la física que subyacen a la vida cotidiana se conocen en su totalidad. Sin duda, ahí fuera podría haber nuevas partículas y fuerzas, pero no interaccionan de forma perceptible con los protones, neutrones y electrones de los que estamos hechos. Y no hace falta decir que queda mucho trabajo por hacer para comprender las reglas que rigen el comportamiento colectivo de muchas de esas partículas. Conocer la teoría del núcleo atómico no nos ayuda mucho con la biología o la psicología, ni

tampoco, de hecho, con la aeronáutica o la climatología. Apenas nos ayuda con la química. Ese es el poder de las teorías efectivas: no necesitamos saber todo lo que ocurre a pequeñas distancias y altas energías para poder hablar con propiedad sobre largas distancias y bajas energías. Pero en el régimen de la vida cotidiana, resulta que conocemos todos los ingredientes y la dinámica relevante de esos ingredientes microscópicos.



La figura superior resume la situación. Las flechas sólidas indican formas de dependencia que están definitivamente ahí, y las flechas discontinuas son posibilidades adicionales. El régimen de la vida cotidiana forma parte de un nivel macroscópico más amplio que incluye el resto del universo, del cual no todo se entiende completamente (ni mucho menos). En el nivel de la teoría cuántica de campos, tenemos la teoría del núcleo atómico y otras posibles partículas y fuerzas. Es muy posible que la astrofísica dependa de todas ellas, pero usted y yo y el mundo con el que interaccionamos solo dependemos de la teoría del núcleo atómico. Y podría haber

niveles aún más profundos, quizá de donde emerja el propio espacio-tiempo. En cierto sentido, también dependemos de lo que ocurra a ese nivel, pero solamente en la medida en que se cruce con la teoría del núcleo atómico. No necesitamos conocer ninguno de los detalles para entender los protones, neutrones y electrones y cómo se unen para formar el mundo que nos rodea.

Hay muchas lagunas posibles en esta línea de razonamiento. Tal vez los dogmas fundamentales de la teoría cuántica de campos estén equivocados de algún modo. Eso sería realmente sorprendente, ya que la TCC es el límite sólido y de baja energía de cualquier teoría basada en la mecánica cuántica, la relatividad y la localidad. Sin embargo, como buenos científicos, debemos admitir que todo es posible. De hecho, la propia mecánica cuántica podría equivocarse de algún modo.

Pero yo no apostaría por ninguna de esas posibilidades. Siempre debemos admitir que no lo sabemos todo, pero no debemos ser tan cautos como para no querer admitir cuando *sí* sabemos cosas. Conocemos las partículas de las que estamos hechos y las fuerzas mediante las cuales interaccionamos. Ese es uno de los mayores logros de la historia de la humanidad. Podemos estar un poquito orgullosos.

---

1 Se suprime, no se prohíbe completamente, porque el Higgs podría convertirse en un par *top/antitop virtual*, siempre que esas partículas virtuales se desintegren rápidamente en otras partículas reales. Las partículas virtuales no obedecen la relación habitual entre masa en reposo y energía. Pero cuanto más se alejan de esa relación habitual, menos probable es que se produzcan, por lo que la desintegración del Higgs a través de pares *top/antitop* sigue siendo muy poco habitual.

2 Los químicos quisquillosos se rasgan las vestiduras ante afirmaciones casuales como esta; el electrón más externo del litio está mucho menos ligado que el del hidrógeno, por lo que existen importantes diferencias químicas entre ambos. Pero ya hemos encolerizado a físicos, matemáticos y filósofos quisquillosos, así que es justo repartir las cosas.

3 El axión es un ejemplo de partícula conjeturada que tiene una masa baja, pero está muy débilmente acoplada. Podría muy bien ser la materia oscura y estar atravesando su cuerpo ahora mismo, pero interacciona tan raramente que no tiene efecto alguno sobre usted. Por eso

los físicos tienen que diseñar enormes y costosos experimentos solo para tener alguna esperanza de detectarla.

APÉNDICE

TRANSFORMADAS DE FOURIER

Los fenómenos ondulatorios plantean un reto matemático especial al físico. El estado clásico de una sola partícula, o de cualquier colección finita de partículas, queda completamente determinado por una lista finita de números: en tres dimensiones, serían las tres componentes de posición y las tres componentes de momento para cada partícula. En cambio, una onda está especificada, en principio, por un número infinito de números: el valor de la onda en cada punto del espacio. En algunos casos especiales podemos tener la suerte de tratar con perfiles de onda que pueden especificarse claramente mediante una función especial concreta, pero, en general, estaremos abrumados de información incluso en el caso más sencillo.

Y el mundo real está lleno de ondas. Incluso desde el punto de vista clásico, el electromagnetismo convenció a los físicos de una vez por todas de la importancia de los campos, y de las ondas dentro de ellos, para describir el mundo. Con la llegada de la mecánica cuántica, la función de onda de lo que desde el punto de vista clásico habría sido una sola partícula pasó a ser una función del espacio y el tiempo. Y pronto nos dimos cuenta de que lo que hay que cuantizar no son partículas, sino campos. La física moderna es una historia de ondas y más ondas.

Podríamos pensar: «Vale, vamos a hacer zoom en un punto del espacio (o del espacio de configuración, en mecánica cuántica) y preguntarnos qué ocurre allí, y luego unir diferentes puntos, al estilo del cálculo». Una idea noble, pero poco práctica. El comportamiento típico de un campo no solamente depende de su valor en un punto concreto, sino también de sus derivadas en el espacio y el tiempo. En

la práctica, esto significa que un campo puede estar haciendo una cosa en un momento dado y otra radicalmente distinta en el siguiente. Todos podemos imaginarnos un mar en calma que se ve atravesado de repente por una ola embravecida.

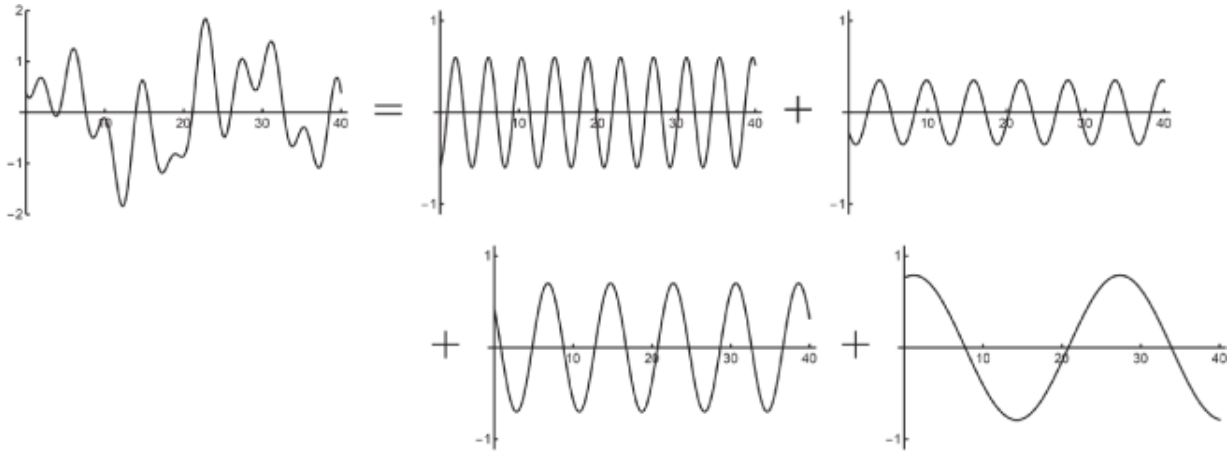
Y sin embargo, de alguna manera, los seres humanos tratamos con ondas todo el tiempo. Si está leyendo este libro, su sistema sensorial y nervioso está recibiendo ondas de luz y convirtiéndolas en información útil en su cerebro. Si está escuchando la versión de audiolibro, ocurre algo parecido, pero con ondas sonoras.

Utilizamos las ondas sonoras y luminosas de diversas maneras, por ejemplo, para determinar con precisión la dirección de la que proceden. Pero centrémonos ahora en nuestra capacidad para reconocer los colores de la luz y los tonos del sonido. En ambos casos, distinguimos la frecuencia o longitud de onda de la onda relevante. Están relacionadas por $f = v/\lambda$, donde f es la frecuencia, λ es la longitud de onda y v es la velocidad de la onda. Lo realmente impresionante es que podemos hacerlo, al menos hasta cierto punto, incluso cuando se superponen un montón de frecuencias diferentes. Lo hacemos cada vez que distinguimos diferentes voces cantando en armonía.

Nuestros cerebros realizan este tipo de procesamiento de forma inconsciente, pero esto plantea una interesante cuestión matemática: dado un perfil de algún tipo de onda, que puede parecer desordenado y nada regular, ¿existe alguna forma sistemática de seleccionar frecuencias específicas?

La respuesta es sí, y la técnica para hacerlo es la **transformada de Fourier**. Debe su nombre al matemático y físico francés Joseph Fourier, que en 1822 publicó un influyente tratado sobre la teoría del flujo de calor entre objetos. En él especulaba que cualquier función puede expresarse como una suma de funciones puramente sinusoidales (tipo seno o coseno), aunque quizá se necesitaría un número infinito de ellas. En la figura se muestra un ejemplo sencillo de una función de aspecto desordenado descompuesta en una suma de cuatro funciones sinusoidales con diferentes amplitudes. Obsérvese que las amplitudes

de las ondas a la derecha de la figura no son todas iguales: la transformada de Fourier cambia información sobre el valor de la función en cada punto por información sobre la amplitud de cada onda en la que se descompone. La idea emocionante es que esto funciona, y la descomposición es perfectamente única y calculable, para cualquier función inicial.



Cuando oí hablar por primera vez de las transformadas de Fourier en la universidad, no tenía ni idea de su utilidad y las consideraba un ejemplo clásico de formalismo matemático sin más. No podía estar más equivocado: en la física moderna, es difícil pensar en una operación más útil o extendida que la transformada de Fourier. Aparecen allí donde hay ondas, es decir, prácticamente en todas partes.

Seamos un poco más específicos. Supongamos que tenemos una función de una sola variable, $f(x)$. Lo que se afirma es que esta función se puede escribir como una suma de funciones sinusoidales (potencialmente, un número infinito de ellas). Cada una de estas funciones sinusoidales va a tener alguna longitud de onda λ. Por diversas razones, es más conveniente utilizar el **número de onda** κ, que es 2π veces el inverso de la longitud de onda:

$$k = 2\pi / \lambda. \qquad \text{(A.1)}$$

Entonces lo que estamos diciendo es que la información contenida en la función original $f(x)$ está contenida equivalentemente en una lista de contribuciones de ondas sinusoidales de todos los números de onda κ posibles. Llamemos a la contribución de una onda concreta $\tilde{f}(k)$. El conjunto de todas estas contribuciones es una función en sí misma, pero es una función de κ, y no de la variable original x. Y la idea básica es que estas dos funciones contienen la misma información:

$$f(x) \leftrightarrow \tilde{f}(k)\,. \tag{A.2}$$

La nueva función se conoce como «transformada de Fourier» de la original. La verdadera complicación que nos encontramos por el camino es que todo es más cómodo si utilizamos números complejos.

Enunciemos directamente las fórmulas para calcular la transformada de Fourier, y también para retroceder hasta la función original:

$$\tilde{f}(k) = \frac{1}{\sqrt{2\pi}} \int f(x) e^{-ikx}\, dx, \tag{A.3a}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int \tilde{f}(k) e^{ikx}\, dk. \tag{A.3b}$$

Observe la diferencia del signo negativo entre las dos exponenciales, y que integramos en x para obtener una función de κ y viceversa. Obsérvese también que estas fórmulas son específicas para funciones unidimensionales; los factores de 2π que van delante serán diferentes para distintos números de dimensiones. Finalmente, diferentes referencias usarán las distintas convenciones de varias maneras. Si está haciendo un cálculo del que depende su vida (o su graduación), asegúrese de comprobar que sus convenciones sean coherentes.

Las exponenciales de los números imaginarios son una buena manera de tratar las funciones sinusoidales, utilizando el siguiente hecho maravilloso:

$$e^{i\theta} = \cos\theta + i\,\mathrm{sen}\,\theta. \qquad \text{(A.4)}$$

Aquí, cosθ y senθ son las conocidas funciones trigonométricas coseno y seno, y e = 2,718... es el número de Euler. El coseno empieza en uno e inicialmente disminuye, mientras que el seno empieza en cero e inicialmente aumenta, oscilando ambos con un periodo 2π. Conocimos la función e^x en *Espacio, tiempo y movimiento* debido a una propiedad especial de su derivada: $\frac{d}{dx}e^x = e^x$. Ahora encontramos otra propiedad divertida de esta función, su relación con la trigonometría cuando el exponente es una variable imaginaria en lugar de real. La ecuación (A.4) se conoce como fórmula de Euler, en honor al polímata suizo Leonhard Euler. Cuando fijamos $\theta = \pi$, obtenemos $e^{i\pi} = \cos\pi + i \sin\pi = -1 + 0$, lo que equivale a la famosa expresión

$$e^{i\pi} + 1 = 0. \qquad \text{(A.5)}$$

Se puede argumentar que 0, 1, i, π y e son los cinco números más importantes de las matemáticas. Esta fórmula relaciona todos ellos con estilo.

¿Qué ocurre realmente en las fórmulas inescrutables (A.3a-b)? En el capítulo 4 señalamos que el espacio de Hilbert, que primero encontramos como el espacio de todas las funciones de onda posibles, puede considerarse un espacio vectorial. Se pueden sumar funciones de onda y multiplicar por números complejos, y el resultado es una función de onda diferente. Pero para una partícula unidimensional con función de onda $\Psi(x)$, el espacio de Hilbert correspondiente es de dimensión infinita. Podemos pensar que cada valor de x representa

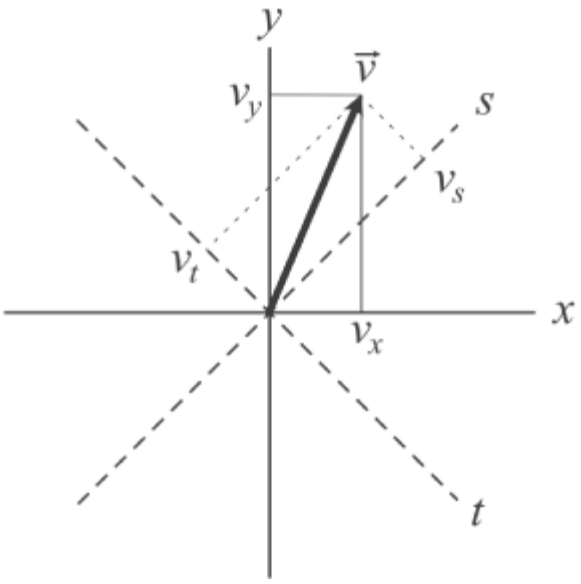
una dimensión diferente del espacio de Hilbert, y el valor $\Psi(x)$ es la componente del vector de dimensión infinita a lo largo de ese eje.

Cuando tiene un espacio vectorial, es libre de cambiar la base que utilice, con el correspondiente cambio de componentes. La figura siguiente muestra un espacio vectorial bidimensional con coordenadas (x, y), en el que también hemos indicado distintas coordenadas (s, t), relacionadas con las originales mediante

$$s = \frac{1}{\sqrt{2}}(x+y), \quad t = \frac{1}{\sqrt{2}}(x-y). \tag{A.6}$$

Los factores de $1/\sqrt{2}$ permiten obtener una transformada inversa elegantemente simétrica, como puede comprobar:

$$x = \frac{1}{\sqrt{2}}(s+t), \quad y = \frac{1}{\sqrt{2}}(s-t). \tag{A.7}$$



Cualquier vector dado $\vec{v}$ puede descomponerse igualmente bien en la base (x, y) o en la base (s, t) como una suma de componentes adecuadas multiplicadas por los vectores de base:

$$\vec{v} = v_x \vec{e}_x + v_y \vec{e}_y = v_s \vec{e}_s + v_t \vec{e}_t. \qquad \text{(A.8)}$$

Las componentes de diferentes bases están relacionadas exactamente igual que las coordenadas. Por ejemplo:

$$v_s = \frac{1}{\sqrt{2}}\left(v_x + v_y\right), \quad v_t = \frac{1}{\sqrt{2}}\left(v_x - v_y\right). \qquad \text{(A.9)}$$

En otras palabras, la información contenida en el vector, codificada en las componentes en una base, puede intercambiarse por las componentes en la otra base, y viceversa:

$$\{v_x, v_y\} \leftrightarrow \{v_s, v_t\}. \qquad \text{(A.10)}$$

Si esto se parece a (A.2), es totalmente intencionado. Una transformada de Fourier no es más que un cambio de base, pero en un espacio vectorial de dimensión infinita. Hablamos de la «base de posiciones» y la «base de momentos», recordando que, en la teoría cuántica de campos, el número de onda y el momento están relacionados por $p = \hbar \kappa$.

Las transformadas de Fourier aparecen en física siempre que resulta útil pensar en las funciones como sumas de ondas con diferentes amplitudes y longitudes de onda, en lugar de como valores en cada punto del espacio. Hay muchas circunstancias de este tipo. Una de ellas es el principio de incertidumbre de Heisenberg, según el cual no existen funciones de onda que estén localizadas simultáneamente en posición y momento. Una vez que sabemos que la función de onda momento-espacio es simplemente la transformada de Fourier de la función de onda posición-espacio, esto tiene mucho sentido. Hemos rotado la base, igual que hicimos en un espacio vectorial bidimensional en la figura anterior. Parece bastante claro que no hay ningún vector que apunte perfectamente en la dirección x (sin componente y) que también apunte perfectamente en las direcciones s

o *t*. Ese es el núcleo del principio de incertidumbre: no se puede estar localizado simultáneamente en términos de dos bases diferentes giradas entre sí. (El contenido no trivial para la física consiste en comprender que la función de onda del momento no es independiente de la función de onda de la posición, sino que es su transformada de Fourier). En general, si una función de x presenta un pico alto, su transformada de Fourier estará muy extendida, y viceversa.

El otro ejemplo obvio es convertir una teoría cuántica de campos libres en una colección de osciladores armónicos simples. Recordemos que existe un término de energía de gradiente en el lagrangiano para un campo, que tiene el aspecto $\frac{1}{2}\,(\partial_x\,\phi)^2$. Si observamos la fórmula (A.3b) para una función de la posición-espacio en términos de su transformada de Fourier, en el lado derecho el único lugar donde aparece x es dentro de la exponencial, $e^{i\kappa x}$. Y su derivada es $\partial_x e^{i\kappa} = i\kappa e^{i\kappa}$. Así que hacer una derivada parcial con respecto a x se puede considerar, una vez que hemos hecho las transformadas de Fourier de todo, como

$$\frac{\partial}{\partial x} \to ik\,. \tag{A.11}$$

Hemos cambiado una derivada por una multiplicación por un número. En lugar de preocuparnos por resolver ecuaciones diferenciales, que pueden ser difíciles, podemos preocuparnos por resolver una ecuación algebraica, que es mucho más fácil.

Físicamente, la conclusión es que un campo escalar libre es una colección de osciladores armónicos simples, uno por cada número de onda, y sabemos cómo resolverlos. Tras la cuantización, los niveles de energía de esos osciladores se convierten en números de partículas. Las transformadas de Fourier explican de dónde proceden los cuantos en la teoría cuántica de campos.